Chi-squared tests for evaluation and comparison of asset pricing models

Nikolay Gospodinov[a,*], Raymond Kan[b], and Cesare Robotti[c]

[a]*Department of Economics, Concordia University, Montreal, Quebec H3G 1M8, Canada, and CIREQ*

[b]*Joseph L. Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada*

[c]*Research Department, Federal Reserve Bank of Atlanta, Atlanta, GA 30309, United States, and EDHEC Risk Institute*

Abstract

This paper presents a general statistical framework for estimation, testing and comparison of asset pricing models using the unconstrained distance measure of Hansen and Jagannathan (1997). The limiting results cover both linear and nonlinear models that could be correctly specified or misspecified. We propose new pivotal specification and model comparison tests that are asymptotically chi-squared distributed. In addition, we develop modified versions of the existing model selection tests with improved finite-sample properties. Finally, we fill an important gap in the literature by providing formal tests of multiple model comparison.

---

*Corresponding author. Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. West, Montreal, Quebec H3G 1M8, Canada, Phone: (514) 848-2424, Fax: (514) 848-4536, E-mail: `nikolay.gospodinov@concordia.ca`.

1. Introduction

It is common for financial economists to view all asset pricing models only as approximations of reality. Although these models are likely to be misspecified, it is still useful to empirically evaluate the degree of misspecification and their relative pricing performance using actual data. In their seminal paper, Hansen and Jagannathan (1997, HJ hereafter) propose measures of model misspecification that are now routinely used for parameter estimation, specification testing and comparison of competing asset pricing models. The unconstrained (constrained) HJ-distance measures the distance between the stochastic discount factor (SDF) of a proposed model and the set of (nonnegative) admissible stochastic discount factors. But despite the recent advances in developing the appropriate econometric theory for comparing asset pricing models based on the HJ-distance, a general statistical procedure for model selection in this context is still incomplete. As a result, many researchers are still ranking alternative models by comparing their corresponding sample HJ-distances without any use of a formal statistical criterion that takes into account the sampling and model misspecification uncertainty. In this paper, we provide a fully-fledged statistical framework for estimation, evaluation and comparison of linear and nonlinear (potentially misspecified) asset pricing models based on the unconstrained HJ-distances. Given some unappealing theoretical properties of the constrained HJ-distance (Gospodinov, Kan and Robotti, 2010a), we do not consider explicitly the sample constrained HJ-distance but the generality of our analytical framework allows us to easily extend the main results for the unconstrained HJ-distance that we derive in this paper to its constrained analog (a detailed econometric analysis of the sample constrained HJ-distance is available from the authors upon request). Our framework could also be used to study the statistical properties of other measures of model misspecification.

The econometric methodology for using the unconstrained HJ-distance as a specification test for linear and nonlinear models is developed by Hansen, Heaton and Luttmer (1995), Jagannathan and Wang (1996) and Parker and Julliard (2005). Kan and Robotti (2009) provide a statistical procedure for comparing linear asset pricing models based on the unconstrained HJ-distance. Furthermore, Kan and Robotti (2009) propose standard errors for the SDF parameter estimates that are valid for misspecified models. The objective of this paper is to provide a unifying framework for improved statistical inference, specification testing and (pairwise and multiple) model comparison based on the sample HJ-distances of competing linear and nonlinear asset pricing models.

Our main contributions can be summarized as follows. First, we propose new Lagrange multiplier tests for individual and joint testing of correct specification of one or more asset pricing models. These new specification tests are asymptotically chi-squared distributed and enjoy improved finite-sample properties compared to the specification test based on the HJ-distance. Second, we derive the non-degenerate joint asymptotic distribution of the parameters and the Lagrange multipliers which are not always asymptotically normally distributed.[1] Third, we improve upon the model selection testing procedures in the existing literature. This is achieved by incorporating the appropriate null hypotheses which leads to simpler model comparison tests that require the estimation of far fewer parameters than the existing testing procedures. While the practice of not imposing the null hypotheses in constructing the test statistics can be justified based on asymptotic arguments, it produces the undesirable outcome of comparing test statistics that are positive by construction (as in the nested model case discussed in Section 3) to distributions that can take on negative values. Our modifications are new to the literature on model selection tests and lead to substantial size and power improvements in setups with many test assets (moment conditions). Importantly, the proposed tests can be easily adapted to other setups including the quasi-likelihood framework of Vuong (1989). Fourth, we propose pivotal (chi-squared) versions of the model comparison tests that are easier to implement and analyze than their weighted chi-squared counterparts. The chi-squared tests appear to possess excellent finite-sample properties and their improved power proves to be particularly important in cases where they are used as pre-tests in sequential testing procedures for strictly non-nested and overlapping models. Fifth, we develop tests for multiple model comparison as well as fast numerical algorithms for computing their asymptotic p-values.[2] Finally, we investigate the finite-sample performance of the proposed inference procedures using Monte Carlo simulations.

The rest of the paper is organized as follows. Section 2 introduces the population and sample HJ-distance problems. It also presents the basic assumptions and the asymptotic properties of the sample HJ-distance and its corresponding estimators. Section 3 develops our pairwise and multiple model comparison tests based on the sample HJ-distances. Section 4 studies the finite-sample properties of our testing procedures using Monte Carlo simulation experiments. Some concluding

[1]This problem is further investigated rigorously in Gospodinov, Kan and Robotti (2010b).

[2]The Matlab codes for implementing all the statistical tests and procedures discussed in the paper are available upon request.

remarks are provided in Section 5. Proofs are collected in the Appendix.

The paper adopts the following notation. Let $\overset{A}{\sim}$ stand for "asymptotically distributed as," χ_p^2 signify a chi-squared random variable with p degrees of freedom, $|w| = (w'w)^{\frac{1}{2}}$ denote the Euclidean norm of a vector w and $||A|| = \sqrt{\text{tr}(A'A)}$ be the Euclidean or Frobenius norm of a matrix A, where $\text{tr}(\cdot)$ is the trace operator. Finally, let $Z = (Z_1, \ldots, Z_s)'$ be a vector of s independent standard normal random variables, and let $\xi = (\xi_1, \ldots, \xi_s)'$ be a vector of s real numbers. Then, $F_s(\xi) = \sum_{i=1}^s \xi_i Z_i^2$ denotes a random variable which is distributed as a weighted sum of s independent chi-squared random variables with one degree of freedom.

2. Hansen-Jagannathan distance

2.1. Population Hansen-Jagannathan distance

Let x_t denote a vector of payoffs of n assets at the end of period t and q_{t-1} be the corresponding costs of these n assets at the end of period $t-1$ with $E[q_{t-1}] \neq 0_n$.[3] This setup can accommodate both gross and excess returns on test assets as well as payoffs of trading strategies that are based on time-varying information. In addition, we assume that $U = E[x_t x_t']$ is nonsingular so that none of the test assets is redundant.

Let m_t represent an admissible SDF at time t and let $\mathcal{M}$ be the set of all admissible SDFs. An SDF m_t is admissible if it prices the test assets correctly, i.e.,[4]

$$E[x_t m_t] = E[q_{t-1}]. \tag{1}$$

Suppose that $y_t(\gamma)$ is a candidate SDF at time t that depends on a k-vector of unknown parameters $\gamma \in \Gamma$, where Γ is the parameter space of γ.[5] An asset pricing model is correctly specified if there exists a $\gamma \in \Gamma$ such that $y_t(\gamma) \in \mathcal{M}$. The model is misspecified if $y_t(\gamma) \notin \mathcal{M}$ for all $\gamma \in \Gamma$. When the asset pricing model is misspecified, we are interested in measuring the degree of model

---

[3]When $E[q_{t-1}] = 0_n$, the mean of the SDF cannot be identified and researchers have to choose some normalization of the SDF (see, for example, Kan and Robotti, 2008).

[4]Strictly speaking, the set of admissible SDFs should be defined in terms of conditional expectations. In this paper, we use an unconditional version of the fundamental pricing equation. This, in principle, could be justified by incorporating conditioning information through scaled payoffs (see, for example, Section 8.1 in Cochrane, 2005).

[5]In this paper, we present results for the case in which the candidate SDF depends on some unknown parameters, but it is straightforward to adapt our analysis to the case in which the SDF does not depend on parameters.

misspecification. HJ suggest using

$$\delta = \min_{\gamma \in \Gamma} \min_{m_t \in \mathcal{M}} \left(E[(y_t(\gamma) - m_t)^2] \right)^{\frac{1}{2}} \tag{2}$$

as a misspecification measure of $y_t(\gamma)$. We refer to δ as the HJ-distance measure.

Instead of solving the above primal problem to obtain δ, HJ suggest that it is sometimes more convenient to solve the following dual problem:

$$\delta^2 = \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} E[y_t(\gamma)^2 - (y_t(\gamma) - \lambda' x_t)^2 - 2\lambda' q_{t-1}], \tag{3}$$

where λ is an n-vector of Lagrange multipliers.

Let $\theta = [\gamma', \lambda']'$ and denote by $\theta^* = [\gamma^{*\prime}, \lambda^{*\prime}]'$ the pseudo-true value that solves the population dual problem in (3):

$$\theta^* = \arg \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} E[\phi_t(\theta)], \tag{4}$$

where

$$\phi_t(\theta) \equiv y_t(\gamma)^2 - m_t(\theta)^2 - 2\lambda' q_{t-1} \tag{5}$$

and

$$m_t(\theta) \equiv y_t(\gamma) - \lambda' x_t. \tag{6}$$

Note that $y_t(\gamma^*)$ prices the n test assets correctly if the vector of pricing errors is zero, i.e.,

$$e(\gamma^*) = E[x_t y_t(\gamma^*) - q_{t-1}] = 0_n. \tag{7}$$

In this case, $y_t(\gamma^*) \in \mathcal{M}$, $\lambda^* = 0_n$ and we refer to γ^* as the true value.[6]

By rearranging the dual problem in (3), it is easy to show that

$$\lambda^* = U^{-1} e(\gamma^*) \tag{8}$$

and

$$\delta^2 = e(\gamma^*)' U^{-1} e(\gamma^*). \tag{9}$$

---

[6]The optimization problem in (4) bears some strong resemblance to the structure of the Euclidean likelihood problem defined as $\min_\gamma \max_\lambda E[\mathbf{h}(\lambda' e(\gamma))]$ with $\mathbf{h}(\varsigma) = -\frac{1}{2}\varsigma^2 - \varsigma$. Other choices of $\mathbf{h}(\varsigma)$ give rise to some popular members of the class of generalized empirical likelihood (GEL) estimators. See Almeida and Garcia (2009) for further discussion of the class of GEL estimators in the context of asset pricing models. While the analysis in this paper can be easily extended to GEL estimators, we choose to present our main results for the HJ-distance measure given its popularity in empirical asset pricing, nice economic (maximum pricing error) interpretation and computational simplicity (closed-form solution for the Lagrange multipliers).

While the quadratic form in the pricing errors in (9) has been widely used in the empirical finance literature for parameter estimation, model evaluation and comparison, the potential usefulness of the information regarding model specification contained in the Lagrange multipliers has been largely ignored. In this paper, we explicitly exploit this information to develop Lagrange multiplier specification tests for individual and multiple models.

2.2. Sample estimators and assumptions

Since the population HJ-distance of a model and its associated parameters are unobservable, they have to be estimated from the data. The estimator of θ^* in (4) is obtained as the solution to the sample dual problem

$$\hat{\theta} = \begin{bmatrix} \hat{\gamma} \\ \hat{\lambda} \end{bmatrix} = \arg \min_{\gamma \in \Gamma} \max_{\lambda \in \Re^n} \frac{1}{T} \sum_{t=1}^{T} \phi_t(\theta). \tag{10}$$

Alternatively, let $e_t(\gamma) = x_t y_t(\gamma) - q_{t-1}$, $e_T(\gamma) = \frac{1}{T} \sum_{t=1}^{T} e_t(\gamma)$ and $\hat{U} = \frac{1}{T} \sum_{t=1}^{T} x_t x_t'$. Then, the estimator $\hat{\theta} = (\hat{\gamma}', \hat{\lambda}')'$ can be obtained sequentially as

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} e_T(\gamma)' \hat{U}^{-1} e_T(\gamma), \tag{11}$$

and

$$\hat{\lambda} = \hat{U}^{-1} e_T(\hat{\gamma}). \tag{12}$$

In the following analysis, we appeal to the empirical process theory to derive the limiting behavior of the estimators and test statistics under correctly specified and misspecified models. The main regularity conditions for the consistency and the asymptotic distribution theory are listed below. They include restrictions on the dependence of the data, identification conditions for the pseudo-true values and some standard assumptions for deriving the limiting distributions.

We first introduce regularity conditions to ensure the stochastic equicontinuity of the sample HJ-distance and the consistency of $\hat{\theta}$.

ASSUMPTION A. *Assume that*

(i) $\phi_t(\theta)$ *is m-dependent,*

(ii) *the parameter space Θ is compact,*

5

(iii) $\phi_t(\theta)$ *is continuous in* $\theta \in \Theta$ *almost surely,*

(iv) $|\phi_t(\theta_1) - \phi_t(\theta_2)| \leq A_t |\theta_1 - \theta_2| \ \forall \ \theta_1, \theta_2 \in \Theta$, *where* A_t *is a bounded random variable that satisfies* $\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E[|A_t|^{2+\omega}] < \infty$ *for some* $\omega > 0$,

(v) $\sup_{\theta \in \Theta} E[|\phi_t(\theta)|^{2+\omega}] < \infty$ *for some* $\omega > 0$,

(vi) *the population dual problem (4) has a unique solution* θ^* *which is in the interior of* Θ.

Assumptions A(i)–A(v) ensure the stochastic equicontinuity of $\phi_t(\theta)$ (see Andrews, 1994 and Stock and Wright, 2000) and imply that

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} \sum_{t=1}^{T} \phi_t(\theta) - E[\phi_t(\theta)] \right| \xrightarrow{p} 0. \tag{13}$$

The m-dependence can be relaxed although results for empirical processes with more general dependence structure are still limited (see, for instance, Andrews, 1993 and Andrews and Pollard, 1994). Assumption A(vi) is an identification condition that ensures the uniqueness of the pseudo-true value θ^*. The uniform convergence in (13) and Assumption A(vi) are sufficient for establishing the consistency of $\hat{\theta}$:

$$\hat{\theta} \xrightarrow{p} \theta^*. \tag{14}$$

Let

$$H \equiv \begin{bmatrix} H_{\gamma\gamma} & H'_{\lambda\gamma} \\ H_{\lambda\gamma} & H_{\lambda\lambda} \end{bmatrix} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 E[\phi_t(\theta^*)]}{\partial\theta\partial\theta'} \tag{15}$$

and

$$M \equiv \begin{bmatrix} M_{\gamma\gamma} & M'_{\lambda\gamma} \\ M_{\lambda\gamma} & M_{\lambda\lambda} \end{bmatrix} = \lim_{T \to \infty} \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial\phi_t(\theta^*)}{\partial\theta} \right]. \tag{16}$$

The next assumption provides conditions for the existence and uniform convergence of the limiting matrices in (15) and (16).

ASSUMPTION B. *Let* $\mathcal{N}(\theta^*)$ *be a neighborhood of* θ^*. *Assume that*

(i) $E[\phi_t(\theta)]$ *is twice continuously differentiable in* θ *for* $\theta \in \mathcal{N}(\theta^*)$,

(ii) $\sup_{\theta \in \mathcal{N}(\theta^*)} \left\| \frac{\partial^2 E[\phi_t(\theta)]}{\partial\theta\partial\theta'} \right\| < \infty$ *and* H *is of full rank,*

(iii) M is a finite positive definite matrix when $\delta > 0$, or $M_{\lambda\lambda}$ is a finite positive definite matrix when $\delta = 0$.

Following Andrews (1994), let $h_t(\theta) = \partial\phi_t(\theta)/\partial\theta$ and define the empirical process $\sqrt{T}\bar{v}_T(\theta)$, where

$$\bar{v}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}v_t(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T}(h_t(\theta) - E[h_t(\theta)]). \tag{17}$$

The next assumption ensures that $\sqrt{T}\bar{v}_T(\theta)$ obeys the central limit theorem.

ASSUMPTION C. *Assume that $v_t(\theta)$ satisfies the conditions*

(i) $|v_t(\theta_1) - v_t(\theta_2)| \le B_t|\theta_1 - \theta_2| \ \forall \ \theta_1, \theta_2 \in \Theta,$ *where B_t is a bounded random variable that satisfies* $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E[|B_t|^{2+\omega}] < \infty$ *for some $\omega > 0$,*

(ii) $\sup_{\theta\in\Theta}E[|v_t(\theta)|^{2+\omega}] < \infty$ *for some $\omega > 0$.*

It proves useful for our subsequent analysis to provide explicit expressions for the partitioned matrices in (15) and (16). Using the fact that

$$\frac{\partial\phi_t(\theta^*)}{\partial\gamma} = 2[y_t(\gamma^*) - m_t(\theta^*)]\frac{\partial y_t(\gamma^*)}{\partial\gamma}, \tag{18}$$

$$\frac{\partial\phi_t(\theta^*)}{\partial\lambda} = 2[x_t m_t(\theta^*) - q_{t-1}], \tag{19}$$

and under Assumptions A, B and C, we can write

$$H_{\gamma\gamma} = 2E\left[(y_t(\gamma^*) - m_t(\theta^*))\frac{\partial^2 y_t(\gamma^*)}{\partial\gamma\partial\gamma'}\right], \tag{20}$$

$$H_{\lambda\gamma} = 2E\left[x_t\frac{\partial y_t(\gamma^*)}{\partial\gamma'}\right], \tag{21}$$

$$H_{\lambda\lambda} = -2E\left[x_t x_t'\right] \equiv -2U, \tag{22}$$

and

$$M_{\lambda\lambda} = 4\sum_{j=-\infty}^{\infty} E\left[(x_t m_t(\theta^*) - q_{t-1})(x_{t+j}m_{t+j}(\theta^*) - q_{t+j-1})'\right]. \tag{23}$$

If the model is correctly specified, we have $\lambda^* = 0_n$ and $y_t(\gamma^*) = m_t(\theta^*)$. Then, it follows that $H_{\gamma\gamma} = 0_{k\times k}$ and $M_{\lambda\lambda} = \sum_{j=-\infty}^{\infty} E[(x_t y_t(\gamma^*) - q_{t-1})(x_{t+j}y_{t+j}(\gamma^*) - q_{t+j-1})']$. Furthermore, we have $\partial\phi_t(\theta^*)/\partial\gamma = 0_k$ which yields $M_{\gamma\gamma} = 0_{k\times k}$ and $M_{\lambda\gamma} = 0_{n\times k}$. This is the reason why Assumption B(iii) requires only $M_{\lambda\lambda}$, and not M, to be positive definite when $\delta = 0$.

7

2.3. Asymptotic results

Let

$$
C = E\left[u_t \frac{\partial^2 y_t(\gamma^*)}{\partial\gamma\partial\gamma'}\right], \tag{24}
$$

$$
D = E\left[x_t \frac{\partial y_t(\gamma^*)}{\partial\gamma'}\right], \tag{25}
$$

$$
S = \sum_{j=-\infty}^{\infty} E\left[e_t(\gamma^*)e_{t+j}(\gamma^*)'\right], \tag{26}
$$

where $u_t = e(\gamma^*)'U^{-1}x_t$.

The following lemma presents the asymptotic distributions of the sample squared HJ-distance under correctly specified and misspecified models.

Lemma 1. *Under Assumptions A, B and C,*

(a) *if* $\delta = 0$,

$$
T\hat{\delta}^2 \overset{A}{\sim} F_{n-k}(\xi), \tag{27}
$$

where the ξ_i's are the eigenvalues of

$$
A = P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P, \tag{28}
$$

with P being an $n \times (n-k)$ orthonormal matrix whose columns are orthogonal to $U^{-\frac{1}{2}}D$.

(b) *if* $\delta > 0$,

$$
\sqrt{T}(\hat{\delta}^2 - \delta^2) \overset{A}{\sim} N(0, \sigma_b^2), \tag{29}
$$

where $\sigma_b^2 = \sum_{j=-\infty}^{\infty} E[b_t b_{t+j}]$ and $b_t = 2u_t y_t - u_t^2 - 2\lambda^{*'}q_{t-1} - \delta^2$.

The asymptotic distribution and matrix A in part (a) of Lemma 1 coincide with the ones derived by Jagannathan and Wang (1996) and Parker and Julliard (2005) for the case of linear and nonlinear models, respectively. The asymptotic normality in part (b) of Lemma 1 has been established by Hansen, Heaton and Luttmer (1995). To conduct inference, the covariance matrices in Lemma 1 should be replaced with consistent estimators. In particular, in part (a), we can replace A with its sample analog

$$
\hat{A} = \hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}, \tag{30}
$$

where $\hat{S}$ is obtained using a nonparametric heteroskedasticity and autocorrelation consistent (HAC) estimator (see, for example, Newey and West, 1987 and Andrews, 1991), $\hat{P}$ is an orthonormal matrix whose columns are orthogonal to $\hat{U}^{-\frac{1}{2}}\hat{D}$ and $\hat{D} = \frac{1}{T}\sum_{t=1}^{T}\left[x_t \frac{\partial y_t(\hat{\gamma})}{\partial \gamma'}\right]$. Similarly, in part (b) we can use a HAC estimator to estimate the variance σ_b^2.

It has been documented (see Ahn and Gadarowski, 2004) that if we use $\hat{A}$ to estimate the eigenvalues ξ_i's, the specification test in part (a) of Lemma 1 tends to overreject substantially when the number of test assets n is large relative to the time series observations T. One way to reduce the overrejection problem is to use a different estimator of S. The consistent estimator of $S_A = M_{\lambda\lambda}/4$, denoted by $\hat{S}_A$, is a good alternative. While $\hat{S}_A$ converges to S under the correctly specified model, $\hat{S}_A$ tends to be larger than $\hat{S}$ in finite samples, thus rendering the overrejection problem less severe.

Lemma 2 below establishes the asymptotic normality of the estimates of the SDF parameters and of the Lagrange multipliers, $\hat{\theta}$, based on the HJ-distance.

Lemma 2. *Under Assumptions A, B and C,*

(a) *if $\delta > 0$,*

$$\sqrt{T}(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_{n+k}, \Sigma), \tag{31}$$

where $\Sigma = \sum_{j=-\infty}^{\infty} E[l_t l'_{t+j}]$ with $l_t = [l'_{1t}, l'_{2t}]'$ given by

$$l_{1t} = (C + D'U^{-1}D)^{-1}\left[D'U^{-1}e_t(\gamma^*) + \left\{\frac{\partial y_t(\gamma^*)}{\partial \gamma} - D'U^{-1}x_t\right\}u_t\right], \tag{32}$$

$$l_{2t} = U^{-1}[Dl_{1t} - e_t(\gamma^*) + x_t u_t]. \tag{33}$$

(b) *if $\delta = 0$,*

$$\sqrt{T}\Pi(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_n, \tilde{\Sigma}), \tag{34}$$

where $\tilde{\Sigma} = \sum_{j=-\infty}^{\infty} E[\tilde{l}_t \tilde{l}'_{t+j}]$ with $\tilde{l}_t = [\tilde{l}'_{1t}, \tilde{l}'_{2t}]'$ given by

$$\tilde{l}_{1t} = (D'U^{-1}D)^{-1}D'U^{-1}e_t(\gamma^*), \tag{35}$$

$$\tilde{l}_{2t} = -P'U^{-\frac{1}{2}}e_t(\gamma^*), \tag{36}$$

and

$$\Pi = \begin{bmatrix} I_k & 0_{k\times n} \\ 0_{(n-k)\times k} & P'U^{\frac{1}{2}} \end{bmatrix}. \tag{37}$$

9

The covariance matrices Σ and $\tilde{\Sigma}$ in Lemma 2 can be consistently estimated using the sample analogs of (32)–(33) and (35)–(36), respectively. Tests of parameter restrictions based on the Wald or distance metric statistics can be easily developed from the results in Lemma 2.

While the estimator $\hat{\gamma}$ is asymptotically normally distributed under both the null and alternative hypotheses, the asymptotic distribution of some linear combinations of $\hat{\lambda}$ is not always normal when $\delta = 0$. To illustrate this, note that when $\delta = 0$, the expression for l_{2t} in (33) simplifies to

$$l_{2t} = [U^{-1}D(D'U^{-1}D)^{-1}D' - I_n]U^{-1}e_t(\gamma^*). \tag{38}$$

Since $D'l_{2t} = 0_k$, the asymptotic covariance matrix of $\sqrt{T}\hat{\lambda}$ is singular when $\delta = 0$. This implies that for a nonzero vector α in the span of the column space of D, $\sqrt{T}\alpha'\hat{\lambda}$ is not asymptotically normal because $\alpha'l_{2t} = 0$.[7]

More generally, Gospodinov, Kan and Robotti (2010b) show that when α is in the span of the column space of D, then

$$T\alpha'\hat{\lambda} \xrightarrow{d} -\tilde{v}_1'v_2, \tag{39}$$

where $\tilde{v}_1$ and v_2 are jointly normally distributed vectors of random variables.

The possible breakdown in the asymptotic normality of $\sqrt{T}\hat{\lambda}$ is the reason why in Lemma 2 we report the asymptotic distribution of $\sqrt{T}P'U^{\frac{1}{2}}\hat{\lambda}$ which always has a non-degenerate asymptotic normal distribution. It is also interesting to note that premultiplying $\hat{\lambda}$ by $P'U^{\frac{1}{2}}$ is similar in spirit to the decomposition of Sowell (1996) in which the n-vector of normalized population moment conditions $U^{-\frac{1}{2}}e_t(\gamma^*)$ is decomposed into k identifying restrictions used for the estimation of γ that characterize the space of identifying restrictions and $(n-k)$ over-identifying restrictions that characterize the space of over-identifying restrictions. This type of decomposition provides the basis for establishing the limiting distribution of the test for over-identifying restrictions. Next, we use the asymptotic result for $\sqrt{T}P'U^{\frac{1}{2}}\hat{\lambda}$ in part (b) of Lemma 2 to develop a Lagrange multiplier (LM) test for model specification.

Theorem 1. *Define the LM statistic as*

$$LM_{\hat{\lambda}} \equiv T\hat{\lambda}'\hat{U}^{\frac{1}{2}}\hat{P}\left(\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\right)^{-1}\hat{P}'\hat{U}^{\frac{1}{2}}\hat{\lambda}. \tag{40}$$

---

[7]It should be emphasized that when the SDF does not have parameters (as in the case of Proposition 4.1 of Hansen, Heaton and Luttmer, 1995), then $\sqrt{T}\hat{\lambda}$ has an asymptotic normal distribution even when $\delta = 0$.

Then, under $H_0 : \delta = 0$ and Assumptions A, B and C,

$$LM_{\hat{\lambda}} \overset{A}{\sim} \chi^2_{n-k}. \tag{41}$$

Since $\delta = 0$ if and only if $\lambda = 0_n$, the LM test in Theorem 1 provides an alternative model specification test that measures the distance of the Lagrange multipliers from zero.[8] Similar arguments can be used for developing an asymptotically equivalent specification test on the model's pricing errors.

3. Model selection tests

In this section, we first present model selection tests of two competing models. Our analysis is similar in spirit to the model selection methodology of Vuong (1989), Rivers and Vuong (2002), Golden (2003), Marcellino and Rossi (2008), and Li, Xu and Zhang (2010), but we provide several improvements upon the results in the literature. First, since for nested models the HJ-distance of the nesting model is always smaller than the HJ-distance of the nested model, the difference between the sample HJ-distances of two nested models should be compared with a distribution that only takes on positive values. However, the existing tests do not impose this restriction and are expected to exhibit size distortions in finite samples. In contrast, we take into account the nested model structure and develop model comparison tests with this desirable property. Second, we develop chi-squared versions of the model comparison tests for strictly non-nested, nested and overlapping models that are easier to implement than the weighted chi-squared tests. In addition to model selection tests of two competing models, we provide multiple model comparison tests that allow us to compare a benchmark model with a set of alternative models in terms of their HJ-distances.

3.1. Pairwise model comparison

Define models

$$\mathcal{F} = \{y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}) \, ; \gamma_{\mathcal{F}} \in \Gamma_{\mathcal{F}}\} \tag{42}$$

---

[8]A similar test, that uses the whole vector of Lagrange multipliers and a generalized inverse of their $n \times n$ asymptotic covariance matrix, is used by Smith (1997) and Imbens, Spady and Johnson (1998) in the context of GEL estimation of moment condition models.

and

$$\mathcal{G} = \{y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}) \, ; \gamma_{\mathcal{G}} \in \Gamma_{\mathcal{G}}\}, \tag{43}$$

where $\gamma_{\mathcal{F}}$ and $\gamma_{\mathcal{G}}$ are k_1 and k_2 parameter vectors, respectively, and $\Gamma_{\mathcal{F}}$ and $\Gamma_{\mathcal{G}}$ denote their corresponding parameter spaces. The population squared HJ-distances for models $\mathcal{F}$ and $\mathcal{G}$ are given by

$$\delta_{\mathcal{F}}^2 = \min_{\gamma_{\mathcal{F}}} \max_{\lambda_{\mathcal{F}}} E[\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}})] \tag{44}$$

$$\delta_{\mathcal{G}}^2 = \min_{\gamma_{\mathcal{G}}} \max_{\lambda_{\mathcal{G}}} E[\phi_t^{\mathcal{G}}(\theta_{\mathcal{G}})], \tag{45}$$

where $\lambda_{\mathcal{F}}$ and $\lambda_{\mathcal{G}}$ are the vectors of Lagrange multipliers for models $\mathcal{F}$ and $\mathcal{G}$, respectively, $\theta_{\mathcal{F}} = [\gamma_{\mathcal{F}}', \, \lambda_{\mathcal{F}}']'$, $\theta_{\mathcal{G}} = [\gamma_{\mathcal{G}}', \, \lambda_{\mathcal{G}}']'$, $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}) \equiv y_t^{\mathcal{F}}(\gamma_{\mathcal{F}})^2 - [m_t^{\mathcal{F}}(\theta_{\mathcal{F}})]^2 - 2\lambda_{\mathcal{F}}' q_{t-1}$, $\phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}) \equiv y_t^{\mathcal{G}}(\gamma_{\mathcal{G}})^2 - [m_t^{\mathcal{G}}(\theta_{\mathcal{G}})]^2 - 2\lambda_{\mathcal{G}}' q_{t-1}$, $m_t^{\mathcal{F}}(\theta_{\mathcal{F}}) \equiv y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}) - \lambda_{\mathcal{F}}' x_t$, and $m_t^{\mathcal{G}}(\theta_{\mathcal{G}}) \equiv y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}) - \lambda_{\mathcal{G}}' x_t$. Denote by $\theta_{\mathcal{F}}^* = [\gamma_{\mathcal{F}}^{*\prime}, \, \lambda_{\mathcal{F}}^{*\prime}]'$ and $\theta_{\mathcal{G}}^* = [\gamma_{\mathcal{G}}^{*\prime}, \, \lambda_{\mathcal{G}}^{*\prime}]'$ the pseudo-true parameters of models $\mathcal{F}$ and $\mathcal{G}$, respectively. If $\mathcal{F} \cap \mathcal{G} = \emptyset$, we have the case of strictly non-nested models. For nested models, we have $\mathcal{F} \subset \mathcal{G}$ or $\mathcal{G} \subset \mathcal{F}$. Finally, if $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, $\mathcal{F} \not\subset \mathcal{G}$, and $\mathcal{G} \not\subset \mathcal{F}$, we refer to $\mathcal{F}$ and $\mathcal{G}$ as overlapping models.

A simple way of testing $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$ is suggested by Hansen, Heaton and Luttmer (1995, pp. 255–256) who establish that the difference between the sample squared HJ-distances of models $\mathcal{F}$ and $\mathcal{G}$ under $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$ is asymptotically normally distributed:

$$\sqrt{T}(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} N(0, \sigma_d^2), \tag{46}$$

where

$$\sigma_d^2 = \sum_{j=-\infty}^{\infty} E[d_t d_{t+j}] \tag{47}$$

and $d_t = \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) - \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$.

It is important to emphasize that the result in (46) holds only if $\sigma_d^2 \neq 0$. To determine whether the use of the normal test in (46) is appropriate, one could do a pre-test of $H_0 : \sigma_d^2 = 0$ (see, for example, Rivers and Vuong, 2002, Golden, 2003 and Marcellino and Rossi, 2008). Alternatively, since $\sigma_d^2 = 0$ if and only if $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$, one could do a pre-test of $H_0 : \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$. This is the approach that we pursue in this paper. [9] There are two possible reasons for $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$:

---

[9]Other inference procedures such as subsampling and *moon* (*m*-out-of- *n*, $m < n$) bootstrap could potentially

(i) the two SDFs are equal, i.e., $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, or (ii) the two SDFs are different but correctly specified, so that $\delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$, which implies $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*) = 0$.

For strictly non-nested models, we cannot have $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. As a result, we only have to test $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ before using the test in (46). [10] For nested models, the test in (46) should not be performed because under $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$, we must have $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{F}}^*)$. The reason is that, in general, the larger model has a smaller HJ-distance and the only case in which the two models can have the same HJ-distance is when $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. Therefore, we should only perform a test of $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ for nested models. Finally, for overlapping models, it is possible that either $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ or $\delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$, so we need to conduct two pre-tests before using the test in (46). We discuss the strictly non-nested, nested and overlapping cases in the following subsections.

3.1.1. Strictly non-nested models

To test $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ for strictly non-nested models, we can use the test statistic $T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)$ based on the difference of the sample HJ-distances of models $\mathcal{F}$ and $\mathcal{G}$. Alternatively, using our results in Section 2, we can also develop an LM test that measures the distance of the Lagrange multipliers of the two models from zero. This will provide a joint test of correct model specification for models $\mathcal{F}$ and $\mathcal{G}$.

To set up the notation, define $e_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = x_t y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) - q_{t-1}$, $e_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = x_t y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) - q_{t-1}$, and

$$\mathcal{S} \equiv \begin{bmatrix} S_{\mathcal{F}} & S_{\mathcal{FG}} \\ S_{\mathcal{GF}} & S_{\mathcal{G}} \end{bmatrix} = \sum_{j=-\infty}^{\infty} E\left[\tilde{e}_t \tilde{e}_{t+j}'\right], \tag{48}$$

where $\tilde{e}_t = [e_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*)', \; e_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)']'$. Let $P_{\mathcal{F}}$ and $P_{\mathcal{G}}$ denote orthonormal matrices with dimensions $n \times (n - k_1)$ and $n \times (n - k_2)$ whose columns are orthogonal to $U^{-\frac{1}{2}} D_{\mathcal{F}}$ and $U^{-\frac{1}{2}} D_{\mathcal{G}}$, respectively, where $D_{\mathcal{F}}$ ($D_{\mathcal{G}}$) is the D matrix for model $\mathcal{F}$ ($\mathcal{G}$) defined in Section 2.3. Also, denote by $\hat{P}_{\mathcal{F}}, \hat{P}_{\mathcal{G}},$

---

deal with the degeneracy of the asymptotic distribution that occurs at $\sigma_d^2 = 0$ and provide correct inference. To the best of our knowledge, the asymptotic validity of these procedures in our context has not been established in the literature. Also, since both of these resampling procedures reduce the number of effective time series observations per moment condition, it is not clear whether they can provide any finite-sample size and power improvements given the excellent finite-sample properties of our asymptotic tests reported in Section 4 below.

[10]In a likelihood framework (see Vuong, 1989), two strictly non-nested models cannot be both correctly specified. However, in our context, a correctly specified model is defined in terms of moment conditions and it is possible for two strictly non-nested models to be both correctly specified. We offer such an example in the Appendix. See Kan and Robotti (2009) and Hall and Pelletier (2011) for further discussion of this point.

$\hat{S}_{\mathcal{F}}$, $\hat{S}_{\mathcal{G}}$, $\hat{S}_{\mathcal{FG}}$, $\hat{S}_{\mathcal{GF}}$, $\hat{\lambda}_{\mathcal{F}}$, and $\hat{\lambda}_{\mathcal{G}}$ the sample counterparts of $P_{\mathcal{F}}$, $P_{\mathcal{G}}$, $S_{\mathcal{F}}$, $S_{\mathcal{G}}$, $S_{\mathcal{FG}}$, $S_{\mathcal{GF}}$, $\lambda_{\mathcal{F}}$ and $\lambda_{\mathcal{G}}$, respectively. Finally, let

$$\tilde{\lambda}_{\mathcal{FG}} = \left[\begin{array}{c} \hat{P}'_{\mathcal{F}} \hat{U}^{\frac{1}{2}} \hat{\lambda}_{\mathcal{F}} \\ \hat{P}'_{\mathcal{G}} \hat{U}^{\frac{1}{2}} \hat{\lambda}_{\mathcal{G}} \end{array} \right]. \tag{49}$$

The following theorem provides the appropriate asymptotic distributions of the difference in the sample squared HJ-distances when both models are correctly specified and an LM test of $H_0 : \lambda_{\mathcal{F}} = \lambda_{\mathcal{G}} = 0_n$ (which is equivalent to testing $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$).

Theorem 2. *Suppose that Assumptions A, B and C hold for each model and $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) \neq y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. Then, under $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$,*

(a)

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{2n-k_1-k_2}(\xi), \tag{50}$$

where the ξ_i's are the eigenvalues of the matrix

$$\left[\begin{array}{cc} P'_{\mathcal{F}} U^{-\frac{1}{2}} S_{\mathcal{F}} U^{-\frac{1}{2}} P_{\mathcal{F}} & -P'_{\mathcal{F}} U^{-\frac{1}{2}} S_{\mathcal{FG}} U^{-\frac{1}{2}} P_{\mathcal{G}} \\ P'_{\mathcal{G}} U^{-\frac{1}{2}} S_{\mathcal{GF}} U^{-\frac{1}{2}} P_{\mathcal{F}} & -P'_{\mathcal{G}} U^{-\frac{1}{2}} S_{\mathcal{G}} U^{-\frac{1}{2}} P_{\mathcal{G}} \end{array} \right], \tag{51}$$

(b)

$$LM_{\tilde{\lambda}_{\mathcal{FG}}} = T\tilde{\lambda}'_{\mathcal{FG}} \hat{\Sigma}_{\tilde{\lambda}_{\mathcal{FG}}}^{-1} \tilde{\lambda}_{\mathcal{FG}} \overset{A}{\sim} \chi^2_{2n-k_1-k_2}, \tag{52}$$

where

$$\hat{\Sigma}_{\tilde{\lambda}_{\mathcal{FG}}} = \left[\begin{array}{cc} \hat{P}'_{\mathcal{F}} \hat{U}^{-\frac{1}{2}} \hat{S}_{\mathcal{F}} \hat{U}^{-\frac{1}{2}} \hat{P}_{\mathcal{F}} & \hat{P}'_{\mathcal{F}} \hat{U}^{-\frac{1}{2}} \hat{S}_{\mathcal{FG}} \hat{U}^{-\frac{1}{2}} \hat{P}_{\mathcal{G}} \\ \hat{P}'_{\mathcal{G}} \hat{U}^{-\frac{1}{2}} \hat{S}_{\mathcal{GF}} \hat{U}^{-\frac{1}{2}} \hat{P}_{\mathcal{F}} & \hat{P}'_{\mathcal{G}} \hat{U}^{-\frac{1}{2}} \hat{S}_{\mathcal{G}} \hat{U}^{-\frac{1}{2}} \hat{P}_{\mathcal{G}} \end{array} \right]. \tag{53}$$

Since the eigenvalues ξ_i's in part (a) of Theorem 2 can take on both positive and negative values, the test of the hypothesis $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ should be two-sided. The LM test in part (b) of Theorem 2 provides an alternative way of testing $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ (using the equivalence between $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ and $H_0 : \lambda_{\mathcal{F}} = \lambda_{\mathcal{G}} = 0_n$) but it is easier to implement and is expected to deliver power gains compared to the test in part (a). The reason is that the test in part (a) may have low power in finite samples when $\hat{\delta}_{\mathcal{F}}^2 \approx \hat{\delta}_{\mathcal{G}}^2 \neq 0$ although it is still consistent since under the alternative $\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2 = O_p(T^{-1/2})$ and $|T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)| \to \infty$.

In summary, our proposed test of equality of the squared HJ-distances of two strictly non-nested models involves first testing whether the two models are both correctly specified using one of the

14

tests in Theorem 2. If we reject, then we can perform the test in (46). Suppose that α_1 and α_2 are the asymptotic significance levels used in the pre-test $\delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ and in the normal test in (46), respectively. Then, our sequential test has a significance level that is asymptotically bounded above by $\max[\alpha_1, \alpha_2]$. Thus, if $\alpha_1 = \alpha_2 = 0.05$, the significance level of this procedure, as a test of $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$, is asymptotically no larger than 5%.

3.1.2. Nested models

For nested models, σ_d^2 is zero by construction under the null of equal HJ-distances. Therefore, the normal test in (46) cannot be used. In addition, for nested models, $\delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$ if and only if $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, so we can simply test $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$.

Let $H_{\mathcal{F}} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 E[\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*)]}{\partial \theta_{\mathcal{F}} \partial \theta_{\mathcal{F}}'}$ and $M_{\mathcal{F}} = \lim_{T \to \infty} \text{Var}\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial \phi_t(\theta_{\mathcal{F}}^*)}{\partial \theta}\right]$ with $H_{\mathcal{G}}$ and $M_{\mathcal{G}}$ defined similarly. Marcellino and Rossi (2008) among others show that under $H_0 : \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$,[11]

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{2n+k_1+k_2}(\xi), \tag{55}$$

where the ξ_i's are the eigenvalues of the matrix

$$\frac{1}{2} \begin{bmatrix} -H_{\mathcal{F}}^{-1} M_{\mathcal{F}} & -H_{\mathcal{F}}^{-1} M_{\mathcal{F}\mathcal{G}} \\ H_{\mathcal{G}}^{-1} M_{\mathcal{G}\mathcal{F}} & H_{\mathcal{G}}^{-1} M_{\mathcal{G}} \end{bmatrix}. \tag{56}$$

Several remarks regarding this inference procedure are in order. First, estimating the ξ_i's from the sample counterpart of the matrix in (56) produces more nonzero estimated ξ_i's than the theory suggests. In addition, the estimated ξ_i's do not have the same sign. This is problematic because for nested models, the larger model has a smaller sample HJ-distance by construction. By not imposing the constraints that the ξ_i's should have the same sign, the nonnegative test statistic $T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)$ is compared with a distribution that can take on both positive and negative values. This could result in serious finite-sample distortions of the test. In the ensuing analysis, we will show that under $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$, some of the ξ_i's are equal to zero and the nonzero ξ_i's have the same sign.

---

[11]Alternatively, we can directly test $H_0 : \sigma_d^2 = 0$. In this case,

$$T\hat{\sigma}_d^2 \overset{A}{\sim} F_{2n+k_1+k_2}(\xi), \tag{54}$$

where $\hat{\sigma}_d^2$ is a consistent estimator of σ_d^2 and the ξ_i's are four times the squared eigenvalues of the matrix in (56) (see Golden, 2003).

Without loss of generality, we assume $\mathcal{F} \subset \mathcal{G}$. Suppose that the null hypothesis $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ can be written as a parametric restriction of the form $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$ for model $\mathcal{G}$ against $H_1 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) \neq 0_{k_2-k_1}$, where $\psi(\cdot)$ is a twice continuously differentiable function in its argument. Define

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{\partial \psi_{\mathcal{G}}(\gamma_{\mathcal{G}})}{\partial \gamma_{\mathcal{G}}'} \tag{57}$$

as a $(k_2 - k_1) \times k_2$ derivative matrix of the parametric restrictions $\psi_{\mathcal{G}}$. For many models of interest, $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}})$ when a subset of the parameters of model $\mathcal{G}$ is equal to zero (or a constant vector c). In this case, we can rearrange the parameters such that $\psi_{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_1)\times k_1}, \ I_{k_2-k_1}]\gamma_{\mathcal{G}} - c$. Then, $\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_1)\times k_1}, \ I_{k_2-k_1}]$, which is a selector matrix that selects only the part of the parameter vector $\gamma_{\mathcal{G}}$ that is not contained in model $\mathcal{F}$. Also, let $\Sigma_{\hat{\gamma}_{\mathcal{G}}}$ be the asymptotic covariance matrix of $\hat{\gamma}_{\mathcal{G}}$ given by the upper left $k \times k$ block of Σ in part (a) of Lemma 2, $\Psi_*^{\mathcal{G}} \equiv \Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, and $\tilde{H}_{\mathcal{G}} = (C^{\mathcal{G}} + D^{\mathcal{G}\prime}U^{-1}D^{\mathcal{G}})^{-1}$, where the matrices C, D, and U are defined in Section 2.3. Finally, define the Wald test statistic

$$Wald_{\hat{\psi}_{\mathcal{G}}} = T\hat{\psi}_{\mathcal{G}}'(\hat{\Psi}^{\mathcal{G}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{G}}}\hat{\Psi}^{\mathcal{G}\prime})^{-1}\hat{\psi}_{\mathcal{G}}, \tag{58}$$

where $\hat{\psi}_{\mathcal{G}} = \psi_{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})$, $\hat{\Psi}^{\mathcal{G}} = \Psi^{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})$, and $\hat{\Sigma}_{\hat{\gamma}_{\mathcal{G}}}$ is a consistent estimator of $\Sigma_{\hat{\gamma}_{\mathcal{G}}}$.

Theorem 3 below presents the asymptotic distribution of $T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)$ and the Wald test under $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$.

Theorem 3. *Suppose that Assumptions A, B and C hold and $\mathcal{F} \subset \mathcal{G}$. Then, under $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$,*

(a)
$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_2-k_1}(\xi), \tag{59}$$

where the ξ_i's are the eigenvalues of the matrix

$$(\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1}\Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}, \tag{60}$$

(b)
$$Wald_{\hat{\psi}_{\mathcal{G}}} \overset{A}{\sim} \chi_{k_2-k_1}^2. \tag{61}$$

16

Part (a) of Theorem 3 shows that, under $H_0: y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$, only $k_2 - k_1$ of the eigenvalues of (56) are nonzero and they all have the same sign.[12] In practice, we need to estimate the ξ_i's to construct the test. Using the sample version of the matrix in part (a) of Theorem 3 instead of the sample version of the matrix in (56) to estimate the ξ_i's results in a substantial reduction of the number of estimated eigenvalues. In addition, the resulting estimated eigenvalues are guaranteed to be positive. The Wald test in part (b) of Theorem 3 offers an alternative way of testing the equality of two nested SDFs by testing directly $H_0: \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2 - k_1}$. This Wald test is asymptotically pivotal and is easier to implement than the test in part (a).

3.1.3.   Overlapping models

For overlapping models, the variance σ_d^2 can be zero when (i) $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ or (ii) both models are correctly specified.[13] Since Theorem 2 is applicable to the second scenario, here we only need to derive the test of $H_0: y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$.

It is well known that for linear models, the equality of the SDFs implies zero restrictions on the parameter vectors (see, for example, Lien and Vuong, 1987 and Kan and Robotti, 2009). Similar restrictions can also be obtained for nonlinear models. Let $y_t^{\mathcal{H}}(\gamma_{\mathcal{H}})$ be the SDF of model $\mathcal{H}$, where $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ and $\gamma_{\mathcal{H}}$ is a k_3-vector. Therefore, $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ implies $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ and $y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$. Suppose that $H_0: y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ and $y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ can be written as a parametric restriction of the form $H_0: \psi_{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = 0_{k_1 - k_3}$ and $\psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2 - k_3}$, where $\psi_{\mathcal{F}}(\cdot)$ and $\psi_{\mathcal{G}}(\cdot)$ are some twice continuously differentiable functions of their arguments. Let

$$\Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}) = \frac{\partial \psi_{\mathcal{F}}(\gamma_{\mathcal{F}})}{\partial \gamma_{\mathcal{F}}'} \tag{63}$$

and

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{\partial \psi_{\mathcal{G}}(\gamma_{\mathcal{G}})}{\partial \gamma_{\mathcal{G}}'} \tag{64}$$

be $(k_1 - k_3) \times k_1$ and $(k_2 - k_3) \times k_2$ derivative matrices of the parametric restrictions $\psi_{\mathcal{F}}$ and $\psi_{\mathcal{G}}$, respectively. In many cases, $H_0: y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ implies that a subset of the parameters of model

---

[12]It can also be shown that for nested models, the test in (54) can be simplified to

$$T\hat{\sigma}_d^2 \overset{A}{\sim} F_{k_2 - k_1}(\xi), \tag{62}$$

where the ξ_i's are four times the squared eigenvalues of the matrix in (60). The proof of this result is available from the authors upon request.

[13]Similar to the case of strictly non-nested models, it is possible for two overlapping SDFs to be both correctly specified. Examples are available upon request.

$\mathcal{F}$ is equal to zero, and $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{G}}^*) = y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ implies that a subset of the parameters of model $\mathcal{G}$ is equal to zero. For such cases, we can arrange the parameters so that $\Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}) = [0_{(k_1-k_3)\times k_3}, \ I_{k_1-k_3}]$ and $\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}) = [0_{(k_2-k_3)\times k_3}, \ I_{k_2-k_3}]$. Let $\Sigma_{\hat{\gamma}_{\mathcal{FG}}}$ be the asymptotic covariance matrix of $\hat{\gamma}_{\mathcal{FG}} = [\hat{\gamma}_{\mathcal{F}}', \ \hat{\gamma}_{\mathcal{G}}']'$, $\tilde{H}_{\mathcal{F}} = (C^{\mathcal{F}} + D^{\mathcal{F}\prime} U^{-1} D^{\mathcal{F}})^{-1}$, $\tilde{H}_{\mathcal{G}} = (C^{\mathcal{G}} + D^{\mathcal{G}\prime} U^{-1} D^{\mathcal{G}})^{-1}$, $\Psi_*^{\mathcal{F}} = \Psi^{\mathcal{F}}(\gamma_{\mathcal{F}}^*)$, $\Psi_*^{\mathcal{G}} = \Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ and

$$\Psi_*^{\mathcal{FG}} \equiv \begin{bmatrix} \Psi_*^{\mathcal{F}} & 0_{(k_1-k_3)\times k_2} \\ 0_{(k_2-k_3)\times k_1} & \Psi_*^{\mathcal{G}} \end{bmatrix}. \tag{65}$$

Define the Wald test statistic

$$Wald_{\hat{\psi}_{\mathcal{FG}}} = T\hat{\psi}_{\mathcal{FG}}'(\hat{\Psi}^{\mathcal{FG}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}\hat{\Psi}^{\mathcal{FG}\prime})^{-1}\hat{\psi}_{\mathcal{FG}}, \tag{66}$$

where $\hat{\psi}_{\mathcal{FG}} = [\psi_{\mathcal{F}}(\hat{\gamma}_{\mathcal{F}})', \ \psi_{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}})']'$,

$$\hat{\Psi}^{\mathcal{FG}} = \begin{bmatrix} \Psi^{\mathcal{F}}(\hat{\gamma}_{\mathcal{F}}) & 0_{(k_1-k_3)\times k_2} \\ 0_{(k_2-k_3)\times k_1} & \Psi^{\mathcal{G}}(\hat{\gamma}_{\mathcal{G}}) \end{bmatrix}, \tag{67}$$

and $\hat{\Sigma}_{\hat{\gamma}_{\mathcal{FG}}}$ is a consistent estimator of $\Sigma_{\hat{\gamma}_{\mathcal{FG}}}$.

The next theorem establishes the asymptotic distribution of $T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2)$ and $Wald_{\hat{\psi}_{\mathcal{FG}}}$ test under the null hypothesis $H_0 : \psi_{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_3}$.

Theorem 4. *Suppose that $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, $\mathcal{F} \not\subset \mathcal{G}$, $\mathcal{G} \not\subset \mathcal{F}$, and Assumptions A, B and C hold. Then, under $H_0 : \psi_{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = 0_{k_1-k_3}$ and $\psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_3}$,*

(a)

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_1+k_2-2k_3}(\xi), \tag{68}$$

where the ξ_i's are the eigenvalues of the matrix

$$\begin{bmatrix} -(\Psi_*^{\mathcal{F}}\tilde{H}_{\mathcal{F}}\Psi_*^{\mathcal{F}\prime})^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1} \end{bmatrix} \Psi_*^{\mathcal{FG}}\Sigma_{\hat{\gamma}_{\mathcal{FG}}}\Psi_*^{\mathcal{FG}\prime}, \tag{69}$$

(b)

$$Wald_{\hat{\psi}_{\mathcal{FG}}} \overset{A}{\sim} \chi^2_{k_1+k_2-2k_3}. \tag{70}$$

Unlike the case of nested models, the eigenvalues in part (a) of Theorem 4 are not always positive because $\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2$ can take on both positive and negative values. As a result, we need to perform a two-sided test of $H_0 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$. Similarly to the nested case, an alternative way

18

of testing the equality of two overlapping SDFs is to directly test the constraints $\psi_{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = 0_{k_1 - k_3}$ and $\psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2 - k_3}$ using the asymptotically pivotal Wald test in part (b) of Theorem 4.

In summary, our proposed sequential testing procedure of equality of the squared HJ-distances of two overlapping models is the following. First, we need to test whether the two models are both correctly specified using either the test in part (a) of Theorem 2 or the chi-squared test in (52). It should be noted that the tests in part (a) of Theorem 2 and part (a) of Theorem 4 are both $O_p(T^{-1})$ and will have low power against each other. Furthermore, the test in part (a) of Theorem 4 will not be consistent against the alternative $H_1 : y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) \neq y_t^{\mathcal{H}}(\gamma_{\mathcal{H}}^*)$ when both models are correctly specified. As a result, our recommendation is to use the LM test in part (b) of Theorem 2 as a pre-test of whether the two models are both correctly specified. If the null is rejected, we can proceed with testing if the SDFs of the two models are equal using the tests in Theorem 4. Finally, if we still reject, we can then perform the normal test in (46). The significance level of this procedure, as a test of $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2$, is asymptotically bounded above by $\max[\alpha_1, \alpha_2, \alpha_3]$, where α_1, α_2, and α_3 are the asymptotic significance levels used in these three tests.

The results in Theorems 3 and 4 offer substantial advantages over the inference procedure (55)–(56) in Section 3.1.2. Imposing the parametric restrictions that directly arise from the structure of the models and the appropriate null hypotheses results in a drastic reduction of the number of weights that are used to compute the critical values of the tests. More specifically, the number of eigenvalues in the weighted chi-squared distribution is reduced from $2n + k_1 + k_2$ to $k_2 - k_1$ for nested and to $k_1 + k_2 - 2k_3$ for overlapping models. This proves to be particularly advantageous when the number of test assets n is large. The reduced dimensions of the matrices in part (a) of Theorems 3 and 4 are expected to lead to improved finite-sample (size and power) behavior of the model selection tests.

3.2. Multiple model comparison

Thus far, we have considered pairwise model comparison. However, when multiple models are involved, pairwise model comparison may not determine unambiguously the best performing model. In this subsection, we develop formal multiple model comparison tests for non-nested and nested models. Our non-nested model comparison test is a multivariate inequality test based on the

results of Wolak (1987, 1989).[14] Suppose we have $p + 1$ models. We are interested in testing the null hypothesis that the benchmark model, model 1 (we could think of model 1 as model $\mathcal{F}$ in the pairwise model comparison subsection), performs at least as well as the other p models. Let δ_i^2 denote the population squared HJ-distance of model i and let $\rho \equiv (\rho_2, \ldots, \rho_{p+1})$, where $\rho_i \equiv \delta_1^2 - \delta_i^2$. Therefore, the null hypothesis is $H_0 : \rho \leq 0_p$ while the alternative is $H_1 : \rho \in \Re^p$.

The test is based on the sample counterpart, $\hat{\rho} \equiv (\hat{\rho}_2, \ldots, \hat{\rho}_{p+1})$, where $\hat{\rho}_i \equiv \hat{\delta}_1^2 - \hat{\delta}_i^2$. Assume that

$$\sqrt{T}(\hat{\rho} - \rho) \overset{A}{\sim} N(0_p, \Omega_{\hat{\rho}}). \tag{71}$$

As in Section 3.1, sufficient conditions for asymptotic normality are: i) $\delta_i^2 > 0$, and ii) the SDFs of the different models are distinct.[15] Let $\tilde{\rho}$ be the optimal solution in the following quadratic programming problem:

$$\min_{\rho}(\hat{\rho} - \rho)'\hat{\Omega}_{\hat{\rho}}^{-1}(\hat{\rho} - \rho) \quad \text{s.t.} \quad \rho \leq 0_p, \tag{72}$$

where $\hat{\Omega}_{\hat{\rho}}$ is a consistent estimator of $\Omega_{\hat{\rho}}$. The likelihood ratio-type test of the null hypothesis is

$$LR = T(\hat{\rho} - \tilde{\rho})'\hat{\Omega}_{\hat{\rho}}^{-1}(\hat{\rho} - \tilde{\rho}). \tag{73}$$

Since the null hypothesis is composite, to construct a test with the desired size, we require the distribution of LR under the least favorable value of ρ, which is $\rho = 0_p$. Under this value, LR follows a "chi-bar-squared distribution,"

$$LR \overset{A}{\sim} \sum_{i=0}^{p} w_{p-i}(\Omega_{\hat{\rho}})X_i, \tag{74}$$

where the X_i's are independent χ^2 random variables with i degrees of freedom, χ_0^2 is simply defined as the constant zero and the weights w_i sum up to one.[16] We use this procedure to obtain asymptotically valid p-values.[17]

---

[14]Kan, Robotti and Shanken (2010) adapt the multivariate inequality test of Wolak (1987, 1989) to compare the performance of alternative asset pricing models in a two-pass cross-sectional regression framework.

[15]As in Section 3.1, pre-tests of correct specification and equality of SDFs can be easily developed also for multiple models by generalizing the chi-squared tests in (52) and (70) to the $p > 1$ case.

[16]An explicit formula for the weights $w_i(\Omega_{\hat{\rho}})$ is given in Kudo (1963) and the computational details are provided in the Appendix. Although the Monte Carlo simulations in the next section show that our asymptotic approximation works well in experiments with realistic sample sizes, researchers could also use bootstrap methods to obtain the p-value for the proposed test statistic. See also Andrews and Soares (2010), among others, for various subsampling, *moon* (m-out-of-n) bootstrap and "plug-in-asymptotic" procedures for parameter inference in models defined by moment inequalities.

[17]There are alternatives to the multivariate inequality test described above. Under the assumption in (71), White

Before using the multivariate inequality test to compare a benchmark model with a set of alternative models, we remove those alternative models that are nested by the benchmark model since, by construction, $\rho_i \leq 0$ in this case. If any of the remaining alternatives is nested by another alternative model, we remove the "nested" model since the δ^2 of the nesting model will be at least as small. Finally, we also remove from consideration any alternative models that nest the benchmark, since the normality assumption on $\hat{\rho}_i$ does not hold under the null hypothesis that $\rho_i = 0$.

Since the multivariate inequality test described above is not applicable when the benchmark is nested by some alternative models, a different multiple model comparison test is needed in this case. When the alternative models nesting the benchmark are nested within each other, we remove the "nested" models since the δ^2 of the nesting model will be at least as small. In this scenario, one could simply use the pairwise model comparison techniques developed in Section 3.1.2. The situation, however, becomes more complicated when the alternative models exhibit an overlapping structure.

Suppose that the benchmark (with k_1 parameter vector γ_1) is nested by model i (with k_i parameter vector γ_i, $i = 2, \ldots, p+1$). Similar to the setup of Section 3.1.2, suppose that $y_t^1(\gamma_1^*) = y_t^i(\gamma_i^*)$ can be written as a parametric restriction of the form $\psi_i(\gamma_i^*) = 0_{k_i-k_1}$, where $\psi_i(\cdot)$ is a twice continuously differentiable function in its argument. The null hypothesis for multiple model comparison can therefore be formulated as $H_0 : \psi_2(\gamma_2^*) = 0_{k_2-k_1}, \ldots, \psi_{p+1}(\gamma_{p+1}^*) = 0_{k_{p+1}-k_1}$. Having derived the asymptotic distribution of $\hat{\gamma}_i$ in Lemma 2, we can use the delta method to obtain the asymptotic distribution of $\hat{\psi}_i = \psi_i(\hat{\gamma}_i)$. Specifically, let

$$\psi = \begin{pmatrix} \psi_2(\gamma_2^*) \\ \vdots \\ \psi_{p+1}(\gamma_{p+1}^*) \end{pmatrix} \tag{75}$$

and denote by $\hat{\psi}$ a consistent estimator of ψ. Also, let $\Sigma_{\hat{\psi}}$ be the asymptotic covariance matrix of $\hat{\psi}$ with rank l under the null hypothesis and $\hat{\Sigma}_{\hat{\psi}}$ denote its consistent estimator. Then, we have

$$Wald_{\hat{\psi}} = T\hat{\psi}'\hat{\Sigma}_{\hat{\psi}}^+\hat{\psi} \overset{A}{\sim} \chi_l^2, \tag{76}$$

where $\hat{\Sigma}_{\hat{\psi}}^+$ is the generalized inverse of $\hat{\Sigma}_{\hat{\psi}}$. To perform this test, we need to determine the rank of

(2000) and Hansen (2005) provide reality check tests that can be used to compare the performance of multiple models. Computing p-values for their tests, however, requires either Monte Carlo simulation or bootstrap methods and can be very time consuming. See Chen and Ludvigson (2009) for an application of the tests of White (2000) and Hansen (2005) to the study of the HJ-distances of competing asset pricing models.

$\Sigma_{\hat{\psi}}$ under the null hypothesis. For linear SDFs, l is simply the number of distinct factors in the set of alternative models minus the number of factors in the benchmark model.[18] For nonlinear SDFs, determining the rank of $\Sigma_{\hat{\psi}}$ under H_0 depends on the particular overlapping structure of the nesting models which needs to be analyzed on a case-by-case basis.

To conclude, if the benchmark model is nested by some competing models, one should separate the set of competing models into two subsets. The first subset will include competing models that nest the benchmark. To test whether the benchmark performs as well as the models in this subset, one can use the chi-squared nested multiple model comparison test described above. The second subset includes competing models that do not nest the benchmark. For this second subset, we can use the test in (74). If we perform each test at a significance level of $\alpha/2$ and fail to reject the null hypothesis in both tests, then, by the Bonferroni inequality, the size of the joint test will be less than or equal to α.

4. Monte Carlo simulations

In this section, we undertake a Monte Carlo experiment to explore the small-sample properties of all the test statistics discussed in the theoretical part of the paper. We focus on linear asset pricing models given their popularity in the literature. To make our simulations more realistic, we calibrate the parameters by using almost 50 years, 1952:2–2000:4, of quarterly gross returns on the three-month T-bill and the well-known 25 Fama-French size and book-to-market portfolios. The time-series sample size is taken to be $T = 120, 240, 360, 480$ and 600. These choices of T reflect sample sizes that are typically encountered in empirical work. The factors and the returns on the test assets are drawn from a multivariate normal distribution (a more detailed description of the various simulation designs can be found in a separate Appendix available on the authors' websites). We compare actual rejection rates over 100,000 iterations to the nominal 10%, 5% and 1% levels of our tests. The parameters in our simulation study are calibrated using actual data from the

---

[18]For example, suppose we have three linear models with factors $[1, f_{1t}]'$, $[1, f_{1t}, f_{2t}, f_{3t}]'$ and $[1, f_{1t}, f_{2t}, f_{4t}]'$, respectively. Note that the first model is the nested (benchmark) model and the last two models are the alternative models nesting the benchmark. The number of distinct factors in the set of alternative models is five $(1, f_{1t}, f_{2t}, f_{3t}$ and $f_{4t})$ and $l = 3$ is obtained by subtracting the number of factors in the benchmark model (1 and f_{1t}). This procedure is used in Section 4.3 below.

following models:[19]

CAPM: the capital asset pricing model

$$y_t^{CAPM} = \gamma_0 + \gamma_1 r_{mkt,t},$$

where r_{mkt} is the excess return on the market portfolio;

CCAPM: the consumption CAPM

$$y_t^{CCAPM} = \gamma_0 + \gamma_1 c_{ndur,t},$$

where c_{ndur} is the log growth rate of non-durable consumption;

YOGO: the durable consumption CAPM of Yogo (2006)

$$y_t^{YOGO} = \gamma_0 + \gamma_1 r_{mkt,t} + \gamma_2 c_{ndur,t} + \gamma_3 c_{dur,t},$$

where c_{dur} denotes the log consumption growth rate of durable goods;

FF3: the three-factor model of Fama and French (1993)

$$y_t^{FF3} = \gamma_0 + \gamma_1 r_{mkt,t} + \gamma_2 r_{smb,t} + \gamma_3 r_{hml,t},$$

where r_{smb} is the return difference between portfolios of small and large stocks and r_{hml} is the return difference between portfolios of high and low book-to-market ratios;

LL: the conditional consumption CAPM of Lettau and Ludvigson (2001)

$$y_t^{LL} = \gamma_0 + \gamma_1 c_{ndur,t} + \gamma_2 cay_{t-1} + \gamma_3 c_{ndur,t} cay_{t-1},$$

where cay is the consumption-wealth ratio.

---

[19]See Gospodinov, Kan and Robotti (2010a) for a detailed description of the various risk factors. In this simulation study, we only consider linear models due to the substantial difficulties involved in generating data from nonlinear asset pricing models that are potentially misspecified with strictly non-nested, nested or overlapping structure. An application of the proposed tests to the comparison of nonlinear asset pricing models, including external and internal habit models, estimated using actual data is available from the authors upon request.

4.1. Parameter estimates and model specification tests

In this subsection, we investigate the size properties of the SDF parameter estimates and the size and power properties of the model specification tests. The data are simulated using two SDFs: CAPM and YOGO. CAPM is an example of models with traded factors since r_{mkt} is a portfolio return. YOGO is an example of models with traded and non-traded factors since c_{ndur} and c_{dur} are macroeconomic factors. As we will see later on, our inference procedures and simulation results are sensitive to whether the included risk factors are traded or not.

We start by analyzing the finite-sample properties of the SDF parameter estimates under model misspecification. One way to summarize the sampling behavior of the SDF parameter estimates and their corresponding asymptotic approximations is to focus on the rejection rates of the t-tests of $H_0 : \gamma_i = 0$. In the simulations, the expected returns are chosen such that the SDF parameter associated with a given factor is equal to zero. The t-tests are constructed using the asymptotic covariance matrices in Lemma 2 and are compared against the critical values from a standard normal distribution. We refer to the t-tests based on (31) and (32) as *t-tests under potentially misspecified models*. For comparison, we also report results using the traditional standard errors derived under correctly specified models based on the asymptotic covariance matrix in (34) and (35). We refer to the corresponding t-tests as *t-tests under correctly specified models*. The reason for investigating the finite-sample performances of the t-tests under correctly specified models in a simulation setup where the model fails to hold exactly is that researchers typically rely on these t-tests in drawing inferences on the SDF parameters even when a model is strongly rejected by the data.

Table 1 presents the empirical size of both t-tests of the null hypothesis $H_0 : \gamma_i = 0$ for realistic values of the HJ-distance measure: $\delta = 0.6524$ for CAPM and $\delta = 0.6514$ for YOGO.[20] Panel A is for the t-tests under potentially misspecified models, while Panel B is for the t-tests under correctly specified models.

$$\boxed{\text{Table 1 about here}}$$

For CAPM, the empirical rejection rates of both t-tests are very close to the nominal size. In

---

[20] To preserve space, we do not report simulation results for the t-ratios associated with the SDF intercept terms.

contrast, the finite-sample performances of these two tests are very different for YOGO. The t-test under potentially misspecified models is well-behaved in this scenario and its empirical size is close to the nominal level with only a slight underrejection.[21] On the contrary, the t-test under correctly specified models tends to overreject substantially. For example, the t-test on the SDF parameter of durable consumption rejects the null hypothesis 30% of the time at the 5% nominal level for $T = 600$. Interestingly, the presence of non-traded factors in YOGO also leads to significant size distortions of the t-test on the traded factor. Finally, in the YOGO case, the performances of the t-tests under correctly specified models deteriorate as T increases.[22]

The difference in behavior between the two t-tests when the model contains non-traded factors warrants some explanation. In the case of linear SDFs, Kan and Robotti (2009) prove that when factors and returns are multivariate elliptically distributed, the standard errors under potentially misspecified models are always bigger than the standard errors constructed under the assumption that the model is correctly specified. They show that the magnitude of the misspecification adjustment term, that reflects the difference between the asymptotic variances of the SDF parameter estimates under correctly specified and misspecified models, depends on, among other things, the degree of model misspecification (as measured by the HJ-distance measure) and the correlations of the factors with the returns. The misspecification adjustment term can be huge when the underlying factor is poorly mimicked by asset returns – a situation that typically arises when some of the factors are macroeconomic variables as in YOGO. Therefore, when the model is misspecified and the factors are poorly spanned by the returns, the t-test under correctly specified models can lead to the erroneous conclusion that certain factors are priced. Our simulation evidence further demonstrates that the t-test under correctly specified models can be seriously oversized and researchers should exercise caution when using it to determine whether a risk factor is priced. Another related issue is the deterioration in the size properties of the t-test under correctly specified models as T increases. This is likely to be a symptom of the fact that some non-traded factors such as c_{ndur} and c_{dur} are almost uncorrelated with the returns. For further discussion, we refer the reader to

---

[21] We should note that the t-test under potentially misspecified models maintains its good size properties even when the data are generated under correctly specified models (results are not reported to conserve space).

[22] The size distortions of the t-tests under correctly specified models documented in Table 1 are somewhat conservative. In unreported simulation experiments, we further analyzed the size properties of the t-tests of $H_0 : \gamma_i = 0$. While the t-test under potentially misspecified models maintained its excellent size properties, there were several instances in which the t-test under correctly specified models exhibited even stronger overrejections. These were typically situations in which the factors and the returns were generated under stronger model misspecification.

Kan and Zhang (1999) who show that when the model is misspecified and a factor is "useless," i.e., independent of the returns, increasing the sample size also increases the severity of the overrejection problem. For these reasons, we strongly recommend using the t-test under potentially misspecified models in factor pricing.

We now turn our attention to the model specification tests developed in Section 2.3. In particular, we assess the finite-sample performance of three specification tests: (i) the HJ-distance test in part (a) of Lemma 1 based on the matrix S, (ii) the HJ-distance test based on the matrix S_A, and (iii) the LM test in Theorem 1. To examine size, we generate returns such that the model holds exactly, i.e., we set $E[x_t] = (1_n - \text{Cov}[x_t, y_t]) / E[y_t]$, where 1_n is an n-vector of ones. The covariance matrix of the factors and returns and the factor means are chosen based on the covariance matrix and the factor means estimated from the data. To examine power, the return means are chosen based on the means estimated from the data, which implies that the population HJ-distances for CAPM and YOGO are 0.6524 and 0.6514, respectively. The empirical size and power of the three tests are presented in Table 2.

$\boxed{\text{Table 2 about here}}$

The overrejections of the HJ-distance test based on S have already been documented in the literature (see, for example, Ahn and Gadarowski, 2004). This typically happens when the number of assets is large relative to the number of time series observations. Our results confirm the overrejections of the HJ-distance test across the different model specifications (24% for CAPM and 17.4% for YOGO at the 5% significance level for $T = 120$) and show that the empirical size approaches the nominal level of the test as T increases (7.3% for CAPM and 6.4% for YOGO at the 5% significance level for $T = 600$).

To the best of our knowledge, the HJ-distance test based on S_A is new to the literature. As we argued in Section 2.3, using $\hat{S}_A$ instead of $\hat{S}$ appears to be particularly important when there are fewer observations per moment condition. This HJ-distance test enjoys much better size properties although it tends to be somewhat conservative. As T increases, the rejection rates approach the nominal size of the test.

Finally, our new LM test has excellent size properties across different models and for all sample sizes. It should be emphasized that the improved sizes of the HJ-distance test that uses S_A and

of the LM test are accompanied by impressive power performance, very similar to the one of the HJ-distance test that uses S. All tables in this section report actual power.[23]

In summary, the LM test in Theorem 1 appears to dominate the other two tests given its simplicity and superior size properties.

4.2.   Model selection tests for strictly non-nested models

In Table 3.A, we evaluate the size and power properties of the tests of $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$. We consider the weighted chi-squared test and the LM test in parts (a) and (b) of Theorem 2, respectively. In light of the discussion in Section 3, one should test whether models $\mathcal{F}$ and $\mathcal{G}$ are correctly specified before applying the normal test in (46).[24]   To examine size, we consider two correctly specified one-factor models (with no intercept) whose factors are generated as r_{mkt} plus two different normal noise terms. To analyze power, we set the return means equal to the means estimated from the sample and compare one of the two one-factor models described above with a model (also with no intercept) that contains a market factor contaminated with a normal noise term, c_{ndur} and c_{dur}. In the size and power experiments, we set the mean of the market factor such that it prices the risk-free asset correctly. This guarantees that the HJ-distances of the two models in the power experiment are close to each other and to values typically encountered in empirical work. The noise terms here and in Section 4.4 have mean zero, standard deviation which is 10% of the standard deviation of r_{mkt} and are independent of the returns and the market factor.

$$\boxed{\text{Table 3 about here}}$$

The size properties of the two tests are quite good when T is bigger than 120. The weighted chi-squared test slightly overrejects for small sample sizes while the LM test slightly underrejects which, again, appears to be due to the relatively large number of test assets considered in the simulation experiment. The empirical size of the tests quickly approaches the nominal level as T increases. Consistent with the discussion below Theorem 2, the LM test delivers nontrivial power

---

[23]Computing size-adjusted power seems infeasible for several of our tests. Their null distributions depend on many nuisance parameters and the simulation of their exact distributions is complicated by the fact that those nuisance parameters are in general not known.

[24]The finite-sample performance of the normal test will be evaluated in Section 4.4 in the context of overlapping models.

gains compared to the weighted chi-squared test and rejects the null $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$ with probability one for $T \geq 240$.

4.3. Model selection tests for nested models

In Panels B and C of Table 3, we report simulation results for pairwise and multiple nested model comparison tests. CCAPM represents our benchmark model. For pairwise model comparison, we consider CCAPM nested by YOGO, while for multiple model comparison we consider CCAPM nested by YOGO and LL. The tests under investigation are the restricted weighted chi-squared test and Wald test in Theorem 3, the unrestricted weighted chi-squared test in (55)–(56) and the chi-squared multiple model comparison test described in Section 3.2.

To analyze the finite-sample behavior of the pairwise model comparison tests under the null of equality of squared HJ-distances, we choose the return means such that the nesting model's slope coefficients associated with the factors that do not belong to the benchmark are zero and both the benchmark and the nesting model are misspecified. To analyze power, the return means are chosen based on the means estimated from the data, which implies that the population HJ-distances for CCAPM and YOGO are 0.6768 and 0.6514, respectively. Turning to multiple model comparison, the size of the chi-squared test is evaluated by choosing the return means such that the nesting models' slope coefficients associated with the factors that do not belong to the benchmark are zero and the benchmark as well as the two nesting models are misspecified. To evaluate power, the return means are chosen based on the means estimated from the data, which implies that the population HJ-distances for CCAPM, YOGO and LL are 0.6768, 0.6514 and 0.6561, respectively.

Table 3.B shows that the test in (55)–(56), that does not impose the restrictions arising from the nested structure of the models, exhibits overrejections that are nontrivial even for $T = 600$ (14.7% and 8.0% at the 10% and 5% nominal levels, respectively). This is due to fact that estimating eigenvalues from the sample counterpart of the matrix in (56) produces more nonzero estimated eigenvalues than the theory suggests. The weighted chi-squared and Wald tests in Theorem 3 have very good size properties and high power (despite the small differences in HJ-distances between models), with the Wald test performing better overall. It should be stressed again that Table 3 reports actual (not size-adjusted) power and the seemingly similar power of the restricted and un-

restricted weighted chi-squared tests is likely due to the overrejections of the unrestricted weighted chi-squared test in (55)–(56) under the null.

For multiple model comparison, the size and power of the chi-squared test in Panel C are impressive. This simulation evidence is very encouraging for the use of this new test in empirical work.

4.4.   Model selection tests for overlapping models

The case of overlapping models is arguably the most important case in practice since many empirical asset pricing specifications contain a constant term and different systematic factors.

Starting with pairwise model comparison, we evaluate the finite-sample behavior of the pre-tests of equality of SDFs in Theorem 4 and the unrestricted weighted chi-squared test in (55)–(56). The simulated data are generated using FF3 and YOGO. To evaluate size, we choose the return means such that the slope coefficients associated with the non-overlapping factors in FF3 and YOGO are zero and the two models are misspecified. To analyze power, the return means are chosen based on the means estimated from the data which implies that the population HJ-distances for FF3 and YOGO are 0.5822 and 0.6514, respectively. Table 3.D shows that all three tests have good size. In terms of power, however, the restricted weighted chi-squared test proposed in part (a) of Theorem 4 performs better than the unrestricted test in (55)–(56) while the chi-squared test in part (b) of Theorem 4 provides further power gains and dominates both weighted chi-squared tests. The high power of the Wald test appears to be particularly important given the fact that this test (along with the test that the models are jointly correctly specified) serves only as a preliminary step in establishing whether two or more models have equal pricing performance.

If the null hypotheses of SDF equality and correct specification of the two models are rejected, then the researcher can proceed with the normal test in (46). In the size computations, the data are simulated from two misspecified three-factor models with intercept, r_{smb} and r_{hml} as common factors and a non-overlapping part that is obtained from contaminating the market factor with two independent normal noise terms defined as in Section 4.2. In the power comparison, the two overlapping models are FF3 and YOGO. Table 3.E ($p = 1$ case) shows that the size properties of the normal test are very good even for small T and that the empirical power quickly approaches 1

as T increases.

Finally, we extend the simulation setup described in the previous paragraph to three three-factor models and employ the LR test in (74) for multiple model comparison. In the size comparison, we add another model with a constant, r_{smb} and r_{hml} as common factors and a non-overlapping part given by the market factor contaminated with independent normal noise. For power evaluation, we consider LL (the benchmark model) in addition to YOGO and FF3. Table 3.E ($p = 2$ case) reveals the very good finite-sample properties of the LR test for comparing multiple asset pricing models.

Overall, these simulation results suggest that the tests developed in this paper should be fairly reliable for the sample sizes typically encountered in empirical work.

5. Concluding remarks

This paper develops a general statistical framework for evaluation and comparison of possibly misspecified asset pricing models using the unconstrained HJ-distance. We propose new pivotal specification and model comparison tests that are asymptotically chi-squared distributed. We also derive new versions of the weighted chi-squared specification and model comparison tests that are computationally efficient and possess improved finite-sample properties. Finally, we develop computationally attractive tests for multiple model comparison. The excellent size and power properties of the proposed tests are demonstrated with simulated data from popular asset pricing models. The simulation results clearly suggest that the standard tests for model specification and selection as well the typical practice of conducting inference on the SDF parameters under the assumption of correctly specified models could be highly misleading in various realistic setups. One of the main findings that emerges from our analysis is that properly incorporating the uncertainty arising from model misspecification as well as imposing the extra restrictions implied by the structure of the models lead to substantially improved inference.

Although our simulation results are encouraging, the small-sample properties of the test statistics proposed in this paper should be explored further. In addition, it is of interest to compare the performance of various linear and nonlinear asset pricing models using the test procedures developed in this paper. A comprehensive empirical application of the proposed methodology is currently being carried out by the authors.

Acknowledgments

Appendix

Preliminary lemma

We first present a preliminary lemma that develops an expansion of the sample HJ-distance that will be used in the proofs of the subsequent lemmas and theorems for model specification and model selection tests.

Lemma A.1. *Under Assumptions A, B and C,*

$$\hat{\delta}^2 - \delta^2 = \frac{1}{T} \sum_{t=1}^{T} (\phi_t(\theta^*) - E[\phi_t(\theta^*)]) - \frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right). \tag{A.1}$$

Proof. We start by expanding $E[\phi_t(\theta^*)] = \delta^2$ about $\hat{\theta}$. Since $\frac{1}{T}\sum_{t=1}^{T} \partial\phi_t(\hat{\theta})/\partial\theta = 0_{n+k}$, we obtain

$$E[\phi_t(\theta^*)] = \frac{1}{T} \sum_{t=1}^{T} \phi_t(\hat{\theta}) - \frac{1}{T} \sum_{t=1}^{T} \left(\phi_t(\hat{\theta}) - E[\phi_t(\hat{\theta})]\right) + \frac{1}{2}(\hat{\theta} - \theta^*)'\frac{\partial^2 E[\phi_t(\tilde{\theta})]}{\partial\theta\partial\theta'}(\hat{\theta} - \theta^*), \tag{A.2}$$

where $\tilde{\theta}$ is an intermediate point between $\hat{\theta}$ and θ^*. Let

$$\bar{h}_T^*(\theta) = \frac{1}{T} \sum_{t=1}^{T} E[h_t(\theta)]. \tag{A.3}$$

A mean value expansion of $\bar{h}_T^*(\theta^*)$ about $\hat{\theta}$ yields

$$0_{n+k} = \sqrt{T}\bar{h}_T^*(\theta^*) = \sqrt{T}\bar{h}_T^*(\hat{\theta}) - \frac{\partial\bar{h}_T^*(\check{\theta})}{\partial\theta}\sqrt{T}(\hat{\theta} - \theta^*), \tag{A.4}$$

where $\check{\theta}$ is another intermediate point on the line segment joining $\hat{\theta}$ and θ^*. From Assumption B(ii) and the consistency of $\hat{\theta}$, we have

$$\sqrt{T}(\hat{\theta} - \theta^*) = H^{-1}\sqrt{T}\bar{h}_T^*(\hat{\theta}) + o_p(1). \tag{A.5}$$

Using the definition of $\bar{v}_T(\theta)$ in (17) and the first order condition of $\frac{1}{T}\sum_{t=1}^{T} h_t(\hat{\theta}) = 0_{n+k}$, it follows that

$$\sqrt{T}\bar{v}_T(\hat{\theta}) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(h_t(\hat{\theta}) - E[h_t(\hat{\theta})]\right) = -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} E[h_t(\hat{\theta})] = -\sqrt{T}\bar{h}_T^*(\hat{\theta}). \tag{A.6}$$

This allows us to rewrite $\sqrt{T}\bar{h}_T^*(\hat{\theta})$ as

$$\sqrt{T}\bar{h}_T^*(\hat{\theta}) = -\sqrt{T}\bar{v}_T(\hat{\theta}) = \sqrt{T}[\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})] - \sqrt{T}\bar{v}_T(\theta^*). \tag{A.7}$$

By the consistency of $\hat{\theta}$, $P[|\hat{\theta} - \theta^*| > \omega] \to 0$ for any arbitrarily small $\omega > 0$. Then,

$$\sqrt{T}|\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})| \leq \sup_{\theta \in \Theta: |\theta - \theta^*| \leq \omega} \sqrt{T}\,|\bar{v}_T(\theta^*) - \bar{v}_T(\theta)|. \tag{A.8}$$

From the stochastic equicontinuity of the empirical process $\sqrt{T}\bar{v}_T(\cdot)$,

$$\sup_{\theta \in \Theta: |\theta - \theta^*| \leq \omega} \sqrt{T}\,|\bar{v}_T(\theta^*) - \bar{v}_T(\theta)| \xrightarrow{p} 0. \tag{A.9}$$

Therefore, we have $\sqrt{T}[\bar{v}_T(\theta^*) - \bar{v}_T(\hat{\theta})] = o_p(1)$ and

$$\sqrt{T}\bar{h}_T^*(\hat{\theta}) = -\sqrt{T}\bar{v}_T(\theta^*) + o_p(1). \tag{A.10}$$

Finally, substituting (A.10) into (A.5) yields

$$\sqrt{T}(\hat{\theta} - \theta^*) = -H^{-1}\sqrt{T}\bar{v}_T(\theta^*) + o_p(1). \tag{A.11}$$

Thus, from (A.11), the consistency of $\hat{\theta}$, and Assumption B(ii), we obtain

$$
\begin{aligned}
\hat{\delta}^2 - \delta^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\hat{\theta}) - E[\phi_t(\hat{\theta})]\right) - \frac{1}{2}(\hat{\theta} - \theta^*)'\frac{\partial^2 E[\phi_t(\tilde{\theta})]}{\partial\theta\partial\theta'}(\hat{\theta} - \theta^*) \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\phi_t(\theta^*) - E[\phi_t(\theta^*)]\right) - \frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right).
\end{aligned} \tag{A.12}
$$

This completes the proof. ■

Proofs

Proof of Lemma 1. (a) From the definition of H in (15), we can use the partitioned matrix inverse formula to obtain

$$H^{-1} = \begin{bmatrix} 2C & 2D' \\ 2D & -2U \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} \tilde{H} & \tilde{H}D'U^{-1} \\ U^{-1}D\tilde{H} & -U^{-1} + U^{-1}D\tilde{H}D'U^{-1} \end{bmatrix}, \tag{A.13}$$

where $\tilde{H} = (C + D'U^{-1}D)^{-1}$. Under the null hypothesis $H_0 : \delta = 0$, (A.1) in Lemma A.1 becomes

$$\hat{\delta}^2 = -\frac{1}{2}\bar{v}_T(\theta^*)'H^{-1}\bar{v}_T(\theta^*) + o_p\left(\frac{1}{T}\right) \tag{A.14}$$

since $\lambda^* = 0_n$ and $\phi_t(\gamma^*, 0_n) = E[\phi_t(\gamma^*, 0_n)] = 0$. Let $\bar{v}_T(\theta^*) = [\bar{v}_{1,T}(\theta^*)', \bar{v}_{2,T}(\theta^*)']'$, where $\bar{v}_{1,T}(\theta^*)$ denotes the first k elements of $\bar{v}_T(\theta^*)$. Under the null, $\bar{v}_{1,T}(\theta^*) = 0_k$ and $C = 0_{k \times k}$.

Then, it follows that

$$
\begin{aligned}
T\hat{\delta}^2 &= -\frac{1}{2}\sqrt{T}\bar{v}_T(\theta^*)'H^{-1}\sqrt{T}\bar{v}_T(\theta^*) + o_p(1) \\
&= \frac{1}{4}\sqrt{T}\bar{v}_{2T}(\theta^*)'[U^{-1} - U^{-1}D(D'U^{-1}D)^{-1}D'U^{-1}]\sqrt{T}\bar{v}_{2,T}(\theta^*) + o_p(1) \\
&= \frac{1}{4}\sqrt{T}\bar{v}_{2T}(\theta^*)'U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}\sqrt{T}\bar{v}_{2,T}(\theta^*) + o_p(1)
\end{aligned}
\tag{A.15}
$$

by using the fact that $I_n - U^{-\frac{1}{2}}D(D'U^{-1}D)^{-1}D'U^{-\frac{1}{2}} = PP'$. Also, Assumptions A, B and C ensure that the empirical process $\sqrt{T}\bar{v}_{2,T}(\theta^*)$ obeys the central limit theorem and

$$
\sqrt{T}\bar{v}_{2,T}(\theta^*) \overset{A}{\sim} N(0_n, M_{\lambda\lambda}).
\tag{A.16}
$$

Thus, using the fact that $M_{\lambda\lambda} = 4S$ under the null, we obtain

$$
T\hat{\delta}^2 \overset{A}{\sim} z'S^{\frac{1}{2}}U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}S^{\frac{1}{2}}z,
\tag{A.17}
$$

where $z \sim N(0_n, I_n)$. Since $S^{\frac{1}{2}}U^{-\frac{1}{2}}PP'U^{-\frac{1}{2}}S^{\frac{1}{2}}$ has the same nonzero eigenvalues as $P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P$, we have

$$
T\hat{\delta}^2 \overset{A}{\sim} F_{n-k}(\xi),
\tag{A.18}
$$

where the ξ_i's are the eigenvalues of $P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P$. This completes the proof of part (a).

(b) Now consider the case $\delta > 0$. In this situation, the asymptotic behavior of $\sqrt{T}(\hat{\delta}^2 - \delta^2)$ is determined by $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\phi_t(\theta^*) - E[\phi_t(\theta^*)])$, which converges weakly to a Gaussian process. Under Assumptions A, B and C, and since $E[\phi_t(\theta^*)] = \delta^2$, we have

$$
\sqrt{T}(\hat{\delta}^2 - \delta^2) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\phi_t(\theta^*) - E[\phi_t(\theta^*)]) + o_p(1) \overset{A}{\sim} N(0, \sigma_b^2).
\tag{A.19}
$$

This completes the proof of part (b). ∎

Proof of Lemma 2. (a) For $\delta > 0$ and under Assumptions A, B and C,

$$
\sqrt{T}\bar{v}_T(\theta^*) \overset{A}{\sim} N(0_{n+k}, M).
\tag{A.20}
$$

Then, combining (A.11) and (A.20), we obtain

$$
\sqrt{T}(\hat{\theta} - \theta^*) \overset{A}{\sim} N(0_{n+k}, H^{-1}MH^{-1}).
\tag{A.21}
$$

To derive an explicit expression for the asymptotic covariance matrix of $\hat{\theta}$, we write

$$H^{-1}MH^{-1} = \sum_{j=-\infty}^{\infty} E[l_t l_{t+j}'], \tag{A.22}$$

where

$$l_t \equiv \begin{bmatrix} l_{1t} \\ l_{2t} \end{bmatrix} = H^{-1} \frac{\partial \phi_t(\theta^*)}{\partial \theta}. \tag{A.23}$$

Using (A.13), (18) and (19), we can express l_{1t} and l_{2t} as

$$l_{1t} = (C + D'U^{-1}D)^{-1} \left[(y_t(\gamma^*) - m_t(\theta^*)) \frac{\partial y_t(\gamma^*)}{\partial \gamma} + D'U^{-1}[x_t m_t(\theta^*) - q_{t-1}] \right], \tag{A.24}$$

$$l_{2t} = U^{-1}[Dl_{1t} - x_t m_t(\theta^*) + q_{t-1}]. \tag{A.25}$$

Using the definition of $m_t(\theta^*)$ and rearranging delivers the desired result. This completes the proof of part (a).

(b) When $\delta = 0$, $C = 0_{k \times k}$ and $m_t(\theta^*) = y_t(\gamma^*)$. Therefore, l_{1t} and l_{2t} simplify to

$$l_{1t} = \tilde{l}_{1t} = (D'U^{-1}D)^{-1}D'U^{-1}e_t(\gamma^*), \tag{A.26}$$

$$l_{2t} = U^{-1}[Dl_{1t} - e_t(\gamma^*)]. \tag{A.27}$$

Premultiplying l_{2t} by $P'U^{\frac{1}{2}}$ yields $\tilde{l}_{2t} = -P'U^{-\frac{1}{2}}e_t(\gamma^*)$. This completes the proof of part (b). ∎

Proof of Theorem 1. From part (b) of Lemma 2, we have

$$\sqrt{T}P'U^{\frac{1}{2}}\hat{\lambda} \overset{A}{\sim} N(0_{n-k}, P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P) \tag{A.28}$$

when $\delta = 0$, or equivalently

$$\sqrt{T}(P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P)^{-\frac{1}{2}}P'U^{\frac{1}{2}}\hat{\lambda} \overset{A}{\sim} N(0_{n-k}, I_{n-k}). \tag{A.29}$$

Then, under Assumptions A, B and C,

$$
\begin{aligned}
LM_{\hat{\lambda}} &= T\hat{\lambda}'\hat{U}^{\frac{1}{2}}\hat{P}\left(\hat{P}'\hat{U}^{-\frac{1}{2}}\hat{S}\hat{U}^{-\frac{1}{2}}\hat{P}\right)^{-1}\hat{P}'\hat{U}^{\frac{1}{2}}\hat{\lambda} \\
&= T\hat{\lambda}'U^{\frac{1}{2}}P\left(P'U^{-\frac{1}{2}}SU^{-\frac{1}{2}}P\right)^{-1}P'U^{\frac{1}{2}}\hat{\lambda} + o_p(1) \\
&\overset{A}{\sim} \chi^2_{n-k}.
\end{aligned}
\tag{A.30}
$$

This completes the proof. ∎

Proof of Theorem 2. (a) From Lemma A.1 and under the null $H_0 : \delta_{\mathcal{F}}^2 = \delta_{\mathcal{G}}^2 = 0$, we can use (A.15) to obtain

$$
\begin{aligned}
&T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \\
&= \frac{1}{4} \left[\begin{array}{c} \sqrt{T}\bar{v}_{2,T}^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ \sqrt{T}\bar{v}_{2,T}^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \end{array} \right]' \left[\begin{array}{cc} U^{-\frac{1}{2}}P_{\mathcal{F}}P_{\mathcal{F}}'U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}}P_{\mathcal{G}}P_{\mathcal{G}}'U^{-\frac{1}{2}} \end{array} \right] \left[\begin{array}{c} \sqrt{T}\bar{v}_{2,T}^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ \sqrt{T}\bar{v}_{2,T}^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \end{array} \right] + o_p(1). \; (\text{A.31})
\end{aligned}
$$

From Assumptions A, B and C, we have

$$
\left[\begin{array}{c} \sqrt{T}\bar{v}_{2,T}^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ \sqrt{T}\bar{v}_{2,T}^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \end{array} \right] \stackrel{A}{\sim} N\left(0_{2n}, 4\mathcal{S}\right). \tag{A.32}
$$

Hence,

$$
T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \stackrel{A}{\sim} z'\mathcal{S}^{\frac{1}{2}} \left[\begin{array}{cc} U^{-\frac{1}{2}}P_{\mathcal{F}}P_{\mathcal{F}}'U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}}P_{\mathcal{G}}P_{\mathcal{G}}'U^{-\frac{1}{2}} \end{array} \right] \mathcal{S}^{\frac{1}{2}}z, \tag{A.33}
$$

where $z \sim N(0_{2n}, I_{2n})$. Furthermore, the nonzero eigenvalues of

$$
\mathcal{S}^{\frac{1}{2}} \left[\begin{array}{cc} U^{-\frac{1}{2}}P_{\mathcal{F}}P_{\mathcal{F}}'U^{-\frac{1}{2}} & 0_{n\times n} \\ 0_{n\times n} & -U^{-\frac{1}{2}}P_{\mathcal{G}}P_{\mathcal{G}}'U^{-\frac{1}{2}} \end{array} \right] \mathcal{S}^{\frac{1}{2}} \tag{A.34}
$$

are the same as the eigenvalues of the matrix

$$
\left[\begin{array}{cc} P_{\mathcal{F}}'U^{-\frac{1}{2}} & 0_{(n-k_1)\times n} \\ 0_{(n-k_2)\times n} & P_{\mathcal{G}}'U^{-\frac{1}{2}} \end{array} \right] \mathcal{S} \left[\begin{array}{cc} U^{-\frac{1}{2}}P_{\mathcal{F}} & 0_{n\times(n-k_2)} \\ 0_{n\times(n-k_1)} & -U^{-\frac{1}{2}}P_{\mathcal{G}} \end{array} \right]. \tag{A.35}
$$

Then, it follows that

$$
T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \stackrel{A}{\sim} F_{2n-k_1-k_2}(\xi), \tag{A.36}
$$

where the ξ_i's are the eigenvalues of the matrix

$$
\left[\begin{array}{cc} P_{\mathcal{F}}'U^{-\frac{1}{2}}S_{\mathcal{F}}U^{-\frac{1}{2}}P_{\mathcal{F}} & -P_{\mathcal{F}}'U^{-\frac{1}{2}}S_{\mathcal{F}\mathcal{G}}U^{-\frac{1}{2}}P_{\mathcal{G}} \\ P_{\mathcal{G}}'U^{-\frac{1}{2}}S_{\mathcal{G}\mathcal{F}}U^{-\frac{1}{2}}P_{\mathcal{F}} & -P_{\mathcal{G}}'U^{-\frac{1}{2}}S_{\mathcal{G}}U^{-\frac{1}{2}}P_{\mathcal{G}} \end{array} \right]. \tag{A.37}
$$

This completes the proof of part (a).

(b) Using the result in part (b) of Lemma 2, it can be shown that when $\lambda_{\mathcal{F}} = \lambda_{\mathcal{G}} = 0_n$,

$$
\sqrt{T}\tilde{\lambda}_{\mathcal{F}\mathcal{G}} \stackrel{A}{\sim} N(0_{2n-k_1-k_2}, \Sigma_{\tilde{\lambda}_{\mathcal{F}\mathcal{G}}}), \tag{A.38}
$$

where

$$
\Sigma_{\tilde{\lambda}_{\mathcal{F}\mathcal{G}}} = \left[\begin{array}{cc} P_{\mathcal{F}}'U^{-\frac{1}{2}}S_{\mathcal{F}}U^{-\frac{1}{2}}P_{\mathcal{F}} & P_{\mathcal{F}}'U^{-\frac{1}{2}}S_{\mathcal{F}\mathcal{G}}U^{-\frac{1}{2}}P_{\mathcal{G}} \\ P_{\mathcal{G}}'U^{-\frac{1}{2}}S_{\mathcal{G}\mathcal{F}}U^{-\frac{1}{2}}P_{\mathcal{F}} & P_{\mathcal{G}}'U^{-\frac{1}{2}}S_{\mathcal{G}}U^{-\frac{1}{2}}P_{\mathcal{G}} \end{array} \right]. \tag{A.39}
$$

Using the fact that $\hat{\Sigma}_{\tilde{\lambda}_{\mathcal{FG}}}$ is a consistent estimator of $\Sigma_{\tilde{\lambda}_{\mathcal{FG}}}$, we have

$$LM_{\tilde{\lambda}_{\mathcal{FG}}} = T\tilde{\lambda}'_{\mathcal{FG}}\hat{\Sigma}^{-1}_{\tilde{\lambda}_{\mathcal{FG}}}\tilde{\lambda}_{\mathcal{FG}} \overset{A}{\sim} \chi^2_{2n-k_1-k_2}. \tag{A.40}$$

This completes the proof of part (b). ∎

Proof of Theorem 3. (a) Since $y_t^{\mathcal{F}}(\gamma_{\mathcal{F}}^*) = y_t^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)$ under the null, it follows that $\lambda_{\mathcal{F}}^* = \lambda_{\mathcal{G}}^*$ and $m_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = m_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$ which implies that $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$. Using these identities, we have

$$\frac{\partial \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*)}{\partial \lambda_{\mathcal{F}}} = 2[x_t m_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) - q_{t-1}] = 2[x_t m_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*) - q_{t-1}] = \frac{\partial \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)}{\partial \lambda_{\mathcal{G}}} \tag{A.41}$$

and

$$\bar{v}_{2,T}^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \bar{v}_{2,T}^{\mathcal{G}}(\theta_{\mathcal{G}}^*). \tag{A.42}$$

It is convenient to express the null hypothesis $H_0 : \psi_{\mathcal{G}}(\gamma_{\mathcal{G}}^*) = 0_{k_2-k_1}$ as a functional dependence

$$H_0 : \gamma_{\mathcal{G}}^* = g(\gamma_{\mathcal{F}}^*), \tag{A.43}$$

where $g(\cdot)$ is a twice continuously differentiable function from $\Gamma_{\mathcal{F}}$ to $\Gamma_{\mathcal{G}}$ (see Gallant, 1987 and Vuong, 1989).[25] Denote by

$$G(\gamma_{\mathcal{F}}) = \frac{\partial g(\gamma_{\mathcal{F}})}{\partial \gamma'_{\mathcal{F}}} \tag{A.44}$$

the $k_2 \times k_1$ matrix of derivatives of $g(\gamma_{\mathcal{F}})$ with respect to $\gamma_{\mathcal{F}}$. Gallant (1987, p. 241) shows that

$$\Psi^{\mathcal{G}}(\gamma_{\mathcal{G}}^*)G(\gamma_{\mathcal{F}}^*) = \Psi^{\mathcal{G}}(g(\gamma_{\mathcal{F}}^*))G(\gamma_{\mathcal{F}}^*) = 0_{(k_2-k_1)\times k_1}. \tag{A.45}$$

Define the matrices

$$\mathbb{S} = [\Psi_*^{\mathcal{G}}, \ 0_{(k_2-k_1)\times n}], \qquad \mathbb{Q} = \begin{bmatrix} G(\gamma_{\mathcal{F}}^*) & 0_{k_2\times n} \\ 0_{n\times k_1} & I_n \end{bmatrix} \tag{A.46}$$

and note that $\mathbb{S}\mathbb{Q} = 0_{(k_2-k_1)\times(n+k_1)}$. Then, using (A.42) and (A.43), it follows that (see Lemma B in Vuong, 1989)

$$\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \mathbb{Q}'\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \tag{A.47}$$

and

$$H_{\mathcal{F}} = \mathbb{Q}'H_{\mathcal{G}}\mathbb{Q}. \tag{A.48}$$

---

[25]Gallant (1987, Section 3.6) provides a discussion of these two alternative representations of the null hypothesis.

By Lemma A.1 and the fact that $\phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*) = \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)$ under the null, we obtain

$$
\begin{aligned}
&T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \\
&= -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)'H_{\mathcal{F}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\
&= -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'\mathbb{Q}(\mathbb{Q}'H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}'\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\
&= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-\frac{1}{2}}[I_{n+k_2} - H_{\mathcal{G}}^{\frac{1}{2}}\mathbb{Q}(\mathbb{Q}'H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}'H_{\mathcal{G}}^{\frac{1}{2}}]H_{\mathcal{G}}^{-\frac{1}{2}}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \quad \text{(A.49)}
\end{aligned}
$$

Using $\mathbb{S}\mathbb{Q} = 0_{(k_2-k_1)\times(n+k_1)}$, it can be shown that (see pp. 241–242 in Gallant, 1987)

$$
I_{n+k_2} - H_{\mathcal{G}}^{\frac{1}{2}}\mathbb{Q}(\mathbb{Q}'H_{\mathcal{G}}\mathbb{Q})^{-1}\mathbb{Q}'H_{\mathcal{G}}^{\frac{1}{2}} = H_{\mathcal{G}}^{-\frac{1}{2}}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-\frac{1}{2}}. \quad \text{(A.50)}
$$

Substituting (A.50) into (A.49) yields

$$
\begin{aligned}
T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) &= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-\frac{1}{2}}[H_{\mathcal{G}}^{-\frac{1}{2}}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-\frac{1}{2}}]H_{\mathcal{G}}^{-\frac{1}{2}}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1) \\
&= \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \quad \text{(A.51)}
\end{aligned}
$$

Furthermore, invoking

$$
\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \overset{A}{\sim} N\left(0_{n+k_2}, M_{\mathcal{G}}\right), \quad \text{(A.52)}
$$

we have

$$
T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} \frac{1}{2}z'\left[M_{\mathcal{G}}^{\frac{1}{2}}H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}M_{\mathcal{G}}^{\frac{1}{2}}\right]z, \quad \text{(A.53)}
$$

where $z \sim N(0_{n+k_2}, I_{n+k_2})$. Denote by $\Sigma_{\hat{\theta}_{\mathcal{G}}}$ the asymptotic covariance matrix of $\hat{\theta}_{\mathcal{G}}$ given in part (a) of Lemma 2. Since the eigenvalues of the matrix $\frac{1}{2}M_{\mathcal{G}}^{\frac{1}{2}}H_{\mathcal{G}}^{-1}\mathbb{S}'(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}M_{\mathcal{G}}^{\frac{1}{2}}$ are the same as the eigenvalues of the matrix

$$
\frac{1}{2}(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}H_{\mathcal{G}}^{-1}M_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}' = \frac{1}{2}(\mathbb{S}H_{\mathcal{G}}^{-1}\mathbb{S}')^{-1}\mathbb{S}\Sigma_{\hat{\theta}_{\mathcal{G}}}\mathbb{S}' = (\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}'})^{-1}\Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}'}, \quad \text{(A.54)}
$$

we conclude that

$$
T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_2-k_1}(\xi), \quad \text{(A.55)}
$$

where the ξ_i's are the eigenvalues of the matrix in (A.54). Since $A = \Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}'}$ and $B = \Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}'}$ are two symmetric positive definite matrices, $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ is also symmetric positive definite with positive eigenvalues. Furthermore, because $A^{-1}B$ and $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ share the same eigenvalues, the eigenvalues of $A^{-1}B$ are also positive. This completes the proof of part (a).

(b) Note that, under the null and using the delta method,

$$\sqrt{T}\hat{\psi}_{\mathcal{G}} \stackrel{A}{\sim} N(0_{k_2-k_1}, \Psi_*^{\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{G}}}\Psi_*^{\mathcal{G}\prime}). \tag{A.56}$$

Substituting consistent estimators for $\Psi_*^{\mathcal{G}}$ and $\Sigma_{\hat{\gamma}_{\mathcal{G}}}$ and constructing the Wald test delivers the desired result. This completes the proof of part (b). ∎

Proof of Theorem 4. (a) Define the following matrices

$$\mathbb{S}_{\mathcal{F}} = \left[\Psi_*^{\mathcal{F}}, \ 0_{(k_1-k_3)\times n}\right], \qquad \mathbb{S}_{\mathcal{G}} = \left[\Psi_*^{\mathcal{G}}, \ 0_{(k_2-k_3)\times n}\right]. \tag{A.57}$$

Since $\mathcal{H} \subset \mathcal{F}$ and $\mathcal{H} \subset \mathcal{G}$, we can use the results from the proof of part (a) of Theorem 3 to obtain

$$T(\hat{\delta}_{\mathcal{H}}^2 - \hat{\delta}_{\mathcal{F}}^2) = \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)'H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}'(\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}')^{-1}\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) + o_p(1) \tag{A.58}$$

and

$$T(\hat{\delta}_{\mathcal{H}}^2 - \hat{\delta}_{\mathcal{G}}^2) = \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}'(\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}')^{-1}\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1). \tag{A.59}$$

Taking the difference yields

$$\begin{aligned}T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) &= -\frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)'H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}'(\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}')^{-1}\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ &\quad + \frac{1}{2}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)'H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}'(\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}')^{-1}\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) + o_p(1).\end{aligned} \tag{A.60}$$

From Assumptions A, B and C, the joint empirical process $\sqrt{T}[\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*)', \bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*)']'$ converges to a Gaussian process:

$$\begin{bmatrix} \sqrt{T}\bar{v}_T^{\mathcal{F}}(\theta_{\mathcal{F}}^*) \\ \sqrt{T}\bar{v}_T^{\mathcal{G}}(\theta_{\mathcal{G}}^*) \end{bmatrix} \stackrel{A}{\sim} N\left(0_{2n+k_1+k_2}, \mathbb{M}\right), \tag{A.61}$$

where

$$\mathbb{M} = \begin{bmatrix} M_{\mathcal{F}} & M_{\mathcal{FG}} \\ M_{\mathcal{GF}} & M_{\mathcal{G}} \end{bmatrix} = \lim_{T\to\infty} \text{Var} \begin{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{\partial \phi_t^{\mathcal{F}}(\theta_{\mathcal{F}}^*)}{\partial \theta_{\mathcal{F}}} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T \frac{\partial \phi_t^{\mathcal{G}}(\theta_{\mathcal{G}}^*)}{\partial \theta_{\mathcal{G}}} \end{bmatrix}. \tag{A.62}$$

Hence,

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \stackrel{A}{\sim}$$
$$z'\left[\frac{1}{2}\mathbb{M}^{\frac{1}{2}}\begin{pmatrix} -H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}'(\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1}\mathbb{S}_{\mathcal{F}}')^{-1}\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}'(\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}_{\mathcal{G}}')^{-1}\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1} \end{pmatrix}\mathbb{M}^{\frac{1}{2}}\right]z, \tag{A.63}$$

39

where $z \sim N(0_{2n+k_1+k_2}, I_{2n+k_1+k_2})$. Then, using the fact that AB and BA share the same nonzero eigenvalues, the matrix in the square brackets in (A.63) has the same nonzero eigenvalues as the matrix

$$\frac{1}{2}\begin{bmatrix} -(\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1}\mathbb{S}'_{\mathcal{F}})^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}'_{\mathcal{G}})^{-1} \end{bmatrix}$$
$$\times \begin{bmatrix} \mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1} & 0_{(k_1-k_3)\times(n+k_2)} \\ 0_{(k_2-k_3)\times(n+k_1)} & \mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1} \end{bmatrix} \mathbb{M} \begin{bmatrix} H_{\mathcal{F}}^{-1}\mathbb{S}'_{\mathcal{F}} & 0_{(n+k_1)\times(k_2-k_3)} \\ 0_{(n+k_2)\times(k_1-k_3)} & H_{\mathcal{G}}^{-1}\mathbb{S}'_{\mathcal{G}} \end{bmatrix}. \quad \text{(A.64)}$$

Using the fact that $\mathbb{S}_{\mathcal{F}}H_{\mathcal{F}}^{-1}\mathbb{S}'_{\mathcal{F}} = \frac{1}{2}\Psi_*^{\mathcal{F}}\tilde{H}_{\mathcal{F}}\Psi_*^{\mathcal{F}\prime}$, $\mathbb{S}_{\mathcal{G}}H_{\mathcal{G}}^{-1}\mathbb{S}'_{\mathcal{G}} = \frac{1}{2}\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime}$ and

$$\Sigma_{\hat{\theta}_{\mathcal{F}\mathcal{G}}} = \begin{bmatrix} H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1} \end{bmatrix} \mathbb{M} \begin{bmatrix} H_{\mathcal{F}}^{-1} & 0_{(n+k_1)\times(n+k_2)} \\ 0_{(n+k_2)\times(n+k_1)} & H_{\mathcal{G}}^{-1} \end{bmatrix} \quad \text{(A.65)}$$

is the asymptotic covariance matrix of $[\hat{\theta}'_{\mathcal{F}}, \hat{\theta}'_{\mathcal{G}}]'$, the matrix in (A.64) can be written as

$$\begin{bmatrix} -(\Psi_*^{\mathcal{F}}\tilde{H}_{\mathcal{F}}\Psi_*^{\mathcal{F}\prime})^{-1} & 0_{(k_1-k_3)\times(k_2-k_3)} \\ 0_{(k_2-k_3)\times(k_1-k_3)} & (\Psi_*^{\mathcal{G}}\tilde{H}_{\mathcal{G}}\Psi_*^{\mathcal{G}\prime})^{-1} \end{bmatrix} \Psi_*^{\mathcal{F}\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}}\Psi_*^{\mathcal{F}\mathcal{G}\prime}. \quad \text{(A.66)}$$

Therefore,

$$T(\hat{\delta}_{\mathcal{F}}^2 - \hat{\delta}_{\mathcal{G}}^2) \overset{A}{\sim} F_{k_1+k_2-2k_3}(\xi), \quad \text{(A.67)}$$

where the ξ_i's are the eigenvalues of the matrix in (A.66). This completes the proof of part (a).

(b) By the delta method,

$$\sqrt{T}\hat{\psi}_{\mathcal{F}\mathcal{G}} \overset{A}{\sim} N(0_{k_1+k_2-2k_3}, \Psi_*^{\mathcal{F}\mathcal{G}}\Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}}\Psi_*^{\mathcal{F}\mathcal{G}\prime}). \quad \text{(A.68)}$$

Using consistent estimators of $\Psi_*^{\mathcal{F}\mathcal{G}}$ and $\Sigma_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}}$ for constructing

$$Wald_{\hat{\psi}_{\mathcal{F}\mathcal{G}}} = T\hat{\psi}'_{\mathcal{F}\mathcal{G}}(\hat{\Psi}^{\mathcal{F}\mathcal{G}}\hat{\Sigma}_{\hat{\gamma}_{\mathcal{F}\mathcal{G}}}\hat{\Psi}^{\mathcal{F}\mathcal{G}\prime})^{-1}\hat{\psi}_{\mathcal{F}\mathcal{G}}, \quad \text{(A.69)}$$

we obtain immediately

$$Wald_{\hat{\psi}_{\mathcal{F}\mathcal{G}}} \overset{A}{\sim} \chi_{k_1+k_2-2k_3}^2. \quad \text{(A.70)}$$

This completes the proof of part (b). ∎

Additional results

Example of two strictly non-nested models that are both correctly specified. Let R be the gross returns on N risky assets and R_0 be the gross return on the risk-free asset. Suppose

40

that R_p is the gross return on the tangency portfolio of the N risky assets. Simple mean-variance mathematics gives

$$E[R] = R_0 1_N + \text{Cov}[R, R_p]\gamma^*, \tag{A.71}$$

where $\gamma^* = (\mu_p - R_0)/\sigma_p^2$, with $\mu_p = E[R_p]$ and $\sigma_p^2 = \text{Var}[R_p]$. In addition, assume that R is multivariate normally distributed. Consider the following SDF

$$y^{\mathcal{F}}(\gamma_{\mathcal{F}}) = \frac{1}{R_0} \exp\left(a_0 - \gamma_{\mathcal{F}} R_p\right), \tag{A.72}$$

where

$$a_0 = -\frac{\gamma^{*2}\sigma_p^2}{2} + \gamma^* \mu_p. \tag{A.73}$$

Let 1_N denote an N-vector of ones. Using Stein's lemma, we can easily establish that

$$E[R_0 y^{\mathcal{F}}(\gamma^*)] = 1, \qquad E[R y^{\mathcal{F}}(\gamma^*)] = 1_N, \tag{A.74}$$

so that $y^{\mathcal{F}}(\gamma^*) \in \mathcal{M}$.

Now consider a factor $f = R_p + \epsilon$, where ϵ is a normal mean-zero measurement error independent of the returns. It follows that $\mu_f = E[f] = \mu_p$ and $\sigma_f^2 = \text{Var}[f] > \sigma_p^2$. Consider an alternative SDF

$$y^{\mathcal{G}}(\gamma_{\mathcal{G}}) = \frac{1}{R_0} \exp\left(\tilde{a}_0 - \gamma_{\mathcal{G}} f\right), \tag{A.75}$$

where

$$\tilde{a}_0 = -\frac{\gamma^{*2}\sigma_f^2}{2} + \gamma^* \mu_p. \tag{A.76}$$

Using Stein's lemma again, we obtain

$$E[R_0 y^{\mathcal{G}}(\gamma^*)] = 1, \qquad E[R y^{\mathcal{G}}(\gamma^*)] = 1_N, \tag{A.77}$$

and $y^{\mathcal{G}}(\gamma^*)$ is also a correctly specified model. Note that $y^{\mathcal{F}}(\gamma_{\mathcal{F}})$ and $y^{\mathcal{G}}(\gamma_{\mathcal{G}})$ are two strictly non-nested models because there are no choices of $\gamma_{\mathcal{F}}$ and $\gamma_{\mathcal{G}}$ that can make these two SDFs identical. This example shows that we can have two strictly non-nested SDFs that are both correctly specified.

Computation of p-values for multivariate inequality test. The biggest hurdle in determining the p-value of LR in (74) is the computation of the weights. For a given $p \times p$ covariance matrix $\Omega = (\omega_{ij})$, the expressions for the weights $w_i(\Omega)$, $i = 0, \ldots, p$, are given in Kudo (1963). The weights

depend on Ω through the correlation coefficients $\rho_{ij} = \omega_{ij}/(\omega_i\omega_j)$. When $p = 1$, $w_0 = w_1 = 1/2$.
When $p = 2$,

$$w_0 = \frac{1}{2} - w_2, \tag{A.78}$$

$$w_1 = \frac{1}{2}, \tag{A.79}$$

$$w_2 = \frac{1}{4} + \frac{\arcsin(\rho_{12})}{2\pi}. \tag{A.80}$$

When $p = 3$,

$$w_0 = \frac{1}{2} - w_2, \tag{A.81}$$

$$w_1 = \frac{1}{2} - w_3, \tag{A.82}$$

$$w_2 = \frac{3}{8} + \frac{\arcsin(\rho_{12\cdot3}) + \arcsin(\rho_{13\cdot2}) + \arcsin(\rho_{23\cdot1})}{4\pi}, \tag{A.83}$$

$$w_3 = \frac{1}{8} + \frac{\arcsin(\rho_{12}) + \arcsin(\rho_{13}) + \arcsin(\rho_{23})}{4\pi}, \tag{A.84}$$

where

$$\rho_{ij\cdot k} = \frac{\rho_{ij} - \rho_{ik}\rho_{jk}}{[(1 - \rho_{ik}^2)(1 - \rho_{jk}^2)]^{\frac{1}{2}}}. \tag{A.85}$$

For $p > 3$, the computation of the weights is more complicated. Following Kudo (1963), let
$\mathcal{P} = \{1, \ldots, p\}$. There are 2^p subsets of $\mathcal{P}$, which are indexed by M. Let $n(M)$ be the number of
elements in M and M' be the complement of M relative to $\mathcal{P}$. Define Ω_M as the submatrix of Ω
that consists of the rows and columns in the set M, $\Omega_{M'}$ as the submatrix of Ω that consists of
the rows and columns in the set M', $\Omega_{M,M'}$ the submatrix of Ω with rows corresponding to the
elements in M and columns corresponding to the elements in M' ($\Omega_{M',M}$ is similarly defined), and
$\Omega_{M\cdot M'} = \Omega_M - \Omega_{M,M'}\Omega_{M'}^{-1}\Omega_{M',M}$. Kudo (1963) shows that

$$w_i(\Omega) = \sum_{M:\, n(M)=i} P(\Omega_{M'}^{-1})P(\Omega_{M\cdot M'}), \tag{A.86}$$

where $P(A)$ is the probability for a multivariate normal distribution with zero mean and covariance
matrix A to have all positive elements. In the above equation, we use the convention that $P[\Omega_{\emptyset\cdot\mathcal{P}}] = 1$ and $P[\Omega_\emptyset^{-1}] = 1$. Using (A.86), we have $w_0(\Omega) = P(\Omega^{-1})$ and $w_p(\Omega) = P(\Omega)$.

Researchers have typically used a Monte Carlo approach to compute the positive orthant prob-
ability $P(A)$. However, the Monte Carlo approach is not efficient because it requires a large number
of simulations to achieve the accuracy of a few digits, even when p is relatively small.

We overcome this problem by using a formula for the positive orthant probability due to Childs (1967) and Sun (1988a). Let $R = (r_{ij})$ be the correlation matrix corresponding to A. Childs (1967) and Sun (1988a) show that

$$
\begin{aligned}
P_{2k}(A) &= \frac{1}{2^{2k}} + \frac{1}{2^{2k-1}\pi} \sum_{1 \le i < j \le 2k} \arcsin(r_{ij}) \\
&\quad + \sum_{j=2}^{k} \frac{1}{2^{2k-j}\pi^j} \sum_{1 \le i_1 < \cdots < i_{2j} \le 2k} I_{2j}\left(R_{(i_1,\ldots,i_{2j})} \right), \qquad (A.87) \\
P_{2k+1}(A) &= \frac{1}{2^{2k+1}} + \frac{1}{2^{2k}\pi} \sum_{1 \le i < j \le 2k+1} \arcsin(r_{ij}) \\
&\quad + \sum_{j=2}^{k} \frac{1}{2^{2k+1-j}\pi^j} \sum_{1 \le i_1 < \cdots < i_{2j} \le 2k+1} I_{2j}\left(R_{(i_1,\ldots,i_{2j})} \right), \qquad (A.88)
\end{aligned}
$$

where $R_{(i_1,\ldots,i_{2j})}$ denotes the submatrix consisting of the $(i_1,\ldots,i_{2j})$-th rows and columns of R, and

$$
I_{2j}(\Lambda) = \frac{(-1)^j}{(2\pi)^j} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\prod_{i=1}^{2j} \frac{1}{\omega_i} \right) \exp\left(-\frac{\omega'\Lambda\omega}{2} \right) d\omega_1 \cdots d\omega_{2j}, \qquad (A.89)
$$

where Λ is a $2j \times 2j$ covariance matrix and $\omega = (\omega_1,\ldots,\omega_{2j})'$. Sun (1988a) provides a recursive relation for $I_{2j}(\Lambda)$ that allows us to obtain I_{2j} starting from I_2. Sun's formula enables us to compute the $2j$-th order multivariate integral I_{2j} using a $(j-1)$-th order multivariate integral, which can be obtained numerically using the Gauss-Legendre quadrature method. Sun (1988b) provides a Fortran subroutine to compute $P(A)$ for $p \le 9$. We improve on Sun's program and are able to accurately compute $P(A)$ and hence $w_i(\Omega)$ for $p \le 11$.

References

Ahn, S. C. and C. Gadarowski, 2004, Small sample properties of the model specification test based on the Hansen-Jagannathan distance. Journal of Empirical Finance 11, 109–132.

Almeida, C. and R. Garcia, 2009, Assessing misspecified asset pricing models with empirical likelihood estimators. Discussion paper, EDHEC Business School.

Andrews, D. W. K., 1991, Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59, 817–858.

Andrews, D. W. K., 1993, An introduction to econometric applications of empirical process theory for dependent random variables. Econometric Reviews 12, 183–216.

Andrews, D. W. K, 1994, Empirical process methods in econometrics, in: R. F. Engle and D. L. Mc-Fadden, (Eds.), Handbook of Econometrics. North-Holland, Amsterdam, pp. 2247–2294.

Andrews, D. W. K. and D. Pollard, 1994, An introduction to functional central limit theorems for dependent stochastic processes. International Statistical Review 62, 119–132.

Andrews, D. W. K. and G. Soares, 2010, Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78, 119–158.

Chen, X. and S. C. Ludvigson, 2009, Land of addicts? An empirical investigation of habit-based asset pricing models. Journal of Applied Econometrics 24, 1057–1093.

Childs, D. R., 1967, Reduction of the multivariate normal integral to characteristic form. Biometrika 54, 293–300.

Cochrane, J. H., 2005, Asset pricing. Princeton University Press, Princeton.

Fama, E. F. and K. R. French, 1993, Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Gallant, R. A., 1987, Nonlinear statistical models. Wiley, New York.

Golden, R. M., 2003, Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. Psychometrika 68, 229–249.

Gospodinov, N., R. Kan and C. Robotti, 2010a, On the Hansen-Jagannathan distance with a no-arbitrage constraint. Discussion paper, Federal Reserve Bank of Atlanta.

Gospodinov, N., R. Kan and C. Robotti, 2010b, Further results on the limiting distribution of GMM sample moment conditions. Discussion paper, Federal Reserve Bank of Atlanta.

Hall, A. R. and D. Pelletier, 2011, Nonnested testing in models estimated via generalized method of moments. Econometric Theory 27, 443–456.

Hansen, L. P., J. C. Heaton and E. G. J. Luttmer, 1995, Econometric evaluation of asset pricing models. Review of Financial Studies 8, 237–274.

Hansen, L. P. and R. Jagannathan, 1997, Assessing specification errors in stochastic discount factor models. Journal of Finance 52, 557–590.

Hansen, P. R., 2005, A test for superior predictive ability. Journal of Business and Economic Statistics 23, 365–380.

Imbens, G. W., R. H. Spady and P. Johnson, 1998, Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

Jagannathan, R. and Z. Wang, 1996, The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3–53.

Kan, R. and C. Robotti, 2008, Specification tests of asset pricing models using excess returns. Journal of Empirical Finance 15, 816–838.

Kan, R. and C. Robotti, 2009, Model comparison using the Hansen-Jagannathan distance. Review of Financial Studies 22, 3449–3490.

Kan, R., C. Robotti and J. Shanken, 2010, Pricing model performance and the two-pass cross-sectional regression methodology. Discussion paper, Federal Reserve Bank of Atlanta.

Kan, R. and C. Zhang, 1999, GMM tests of stochastic discount factor models with useless factors. Journal of Financial Economics 54, 103–127.

Kudo, A., 1963, A multivariate analogue of the one-sided test. Biometrika 50, 403–418.

Lettau, M. and S. C. Ludvigson, 2001, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying. Journal of Political Economy 109, 1238–1287.

Li, H., Y. Xu and X. Zhang, 2010, Evaluating asset pricing models using the second Hansen-Jagannathan distance. Journal of Financial Economics 97, 279–301.

Lien, D. and Q. H. Vuong, 1987, Selecting the best linear regression model: A classical approach. Journal of Econometrics 35, 3–23.

Marcellino, M. and B. Rossi, 2008, Model selection for nested and overlapping nonlinear, dynamic and possibly mis-specified models. Oxford Bulletin of Economics and Statistics 70, 867–893.

Newey, W. K. and K. D. West, 1987, A simple positive semi-definite heteroskedasticity and auto-correlation consistent covariance matrix estimator. Econometrica 55, 703–708.

Parker, J. A. and C. Julliard, 2005, Consumption risk and the cross section of expected returns. Journal of Political Economy 113, 185–222.

Rivers, D. and Q. H. Vuong, 2002, Model selection tests for nonlinear dynamic models. Econometrics Journal 5, 1–39.

Smith, R. J., 1997, Alternative semi-parametric likelihood approaches to generalized method of moments estimation. Economic Journal 107, 503–519.

Sowell, F. B., 1996, Optimal tests for parameter instability in the generalized method of moments framework. Econometrica 64, 1085–1107.

Stock, J. H. and J. H. Wright, 2000, GMM with weak identification. Econometrica 68, 1055–1096.

Sun, H., 1988a, A general reduction method for n-variate normal orthant probability. Communications in Statistics – Theory and Methods 11, 3913–3921.

Sun, H., 1988b, A Fortran subroutine for computing normal orthant probabilities of dimensions up to nine. Communications in Statistics – Simulation and Computation 17, 1097–1111.

Vuong, Q. H., 1989, Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

White, H., 2000, A reality check for data snooping. Econometrica 68, 1097–1126.

Wolak, F. A., 1987, An exact test for multiple inequality and equality constraints in the linear regression model. Journal of the American Statistical Association 82, 782–793.

Wolak, F. A., 1989, Testing inequality constraints in linear econometric models. Journal of Econometrics 31, 205–235.

Yogo, M., 2006, A consumption-based explanation of expected stock returns. Journal of Finance 61, 539–580.

Table 1

t-tests under model misspecification

Panel A: *t*-tests under potentially misspecified models

| | CAPM | | | YOGO | | | | | | | | |
| | Size (*mkt*) | | | Size (*mkt*) | | | Size (*ndur*) | | | Size (*dur*) | | |
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.097	0.047	0.008	0.076	0.032	0.004	0.083	0.037	0.006	0.084	0.039	0.006
240	0.098	0.048	0.009	0.078	0.033	0.004	0.085	0.039	0.006	0.083	0.038	0.006
360	0.097	0.048	0.009	0.082	0.037	0.005	0.089	0.042	0.006	0.087	0.040	0.006
480	0.098	0.049	0.010	0.085	0.039	0.006	0.090	0.044	0.007	0.088	0.041	0.007
600	0.098	0.049	0.010	0.089	0.041	0.006	0.092	0.044	0.008	0.091	0.042	0.007

Panel B: *t*-tests under correctly specified models

| | CAPM | | | YOGO | | | | | | | | |
| | Size (*mkt*) | | | Size (*mkt*) | | | Size (*ndur*) | | | Size (*dur*) | | |
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.098	0.048	0.008	0.157	0.086	0.018	0.282	0.189	0.066	0.291	0.199	0.073
240	0.098	0.048	0.009	0.194	0.117	0.033	0.317	0.225	0.097	0.331	0.239	0.106
360	0.098	0.049	0.009	0.221	0.140	0.046	0.337	0.248	0.120	0.357	0.265	0.130
480	0.099	0.049	0.010	0.236	0.153	0.057	0.352	0.263	0.132	0.374	0.283	0.147
600	0.099	0.049	0.010	0.248	0.166	0.064	0.360	0.272	0.142	0.385	0.296	0.158

The table presents the empirical size of the *t*-tests of $H_0 : \gamma_i = 0$. We report results for different levels of significance (10%, 5% and 1% levels) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the factors and the returns are generated from a multivariate normal distribution. The various *t*-ratios are compared to the critical values from a standard normal distribution. Panel A reports results for *t*-tests under potentially misspecified models based on the asymptotic covariance matrix in (31) and (32), while Panel B reports results for *t*-tests under correctly specified models based on the asymptotic covariance matrix in (34) and (35).

Table 2

Model specification tests

Panel A: HJ-distance test using S

	CAPM						YOGO					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.343	0.240	0.101	1.000	0.999	0.995	0.267	0.174	0.064	0.991	0.982	0.946
240	0.202	0.123	0.037	1.000	1.000	1.000	0.166	0.096	0.027	1.000	1.000	0.999
360	0.162	0.092	0.025	1.000	1.000	1.000	0.139	0.077	0.019	1.000	1.000	1.000
480	0.145	0.079	0.020	1.000	1.000	1.000	0.127	0.068	0.016	1.000	1.000	1.000
600	0.134	0.073	0.017	1.000	1.000	1.000	0.122	0.064	0.014	1.000	1.000	1.000

Panel B: HJ-distance test using S_A

	CAPM						YOGO					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.061	0.019	0.001	0.988	0.958	0.716	0.052	0.017	0.001	0.941	0.869	0.563
240	0.083	0.036	0.005	1.000	1.000	1.000	0.072	0.030	0.004	1.000	0.999	0.995
360	0.089	0.041	0.006	1.000	1.000	1.000	0.080	0.036	0.005	1.000	1.000	1.000
480	0.091	0.042	0.007	1.000	1.000	1.000	0.084	0.039	0.006	1.000	1.000	1.000
600	0.093	0.044	0.008	1.000	1.000	1.000	0.086	0.040	0.007	1.000	1.000	1.000

Panel C: LM test

	CAPM						YOGO					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.117	0.054	0.008	0.995	0.984	0.905	0.121	0.058	0.009	0.982	0.959	0.846
240	0.111	0.054	0.010	1.000	1.000	1.000	0.104	0.050	0.009	1.000	1.000	0.999
360	0.106	0.053	0.010	1.000	1.000	1.000	0.099	0.049	0.009	1.000	1.000	1.000
480	0.104	0.051	0.010	1.000	1.000	1.000	0.099	0.049	0.009	1.000	1.000	1.000
600	0.103	0.052	0.010	1.000	1.000	1.000	0.099	0.047	0.009	1.000	1.000	1.000

The table presents the empirical size and power of three tests of $H_0 : \delta^2 = 0$. Panel A is for the test in part (a) of Lemma 1 that uses the matrix S. Panel B is for the HJ-distance test based on S_A. Finally, Panel C is for the LM test in Theorem 1. We report results for different levels of significance (10%, 5% and 1% levels) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the factors and the returns are generated from a multivariate normal distribution.

Table 3

Model selection tests

Panel A: Joint tests of correct specification for two strictly non-nested SDFs

	Weighted χ^2 test						LM test					
	Size			Power			Size			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.187	0.113	0.035	0.585	0.499	0.336	0.049	0.013	0.000	0.831	0.663	0.247
240	0.138	0.077	0.020	0.806	0.747	0.611	0.078	0.031	0.004	1.000	1.000	1.000
360	0.127	0.069	0.016	0.913	0.881	0.791	0.088	0.038	0.005	1.000	1.000	1.000
480	0.119	0.063	0.014	0.961	0.944	0.892	0.092	0.042	0.006	1.000	1.000	1.000
600	0.115	0.060	0.013	0.984	0.975	0.946	0.093	0.043	0.007	1.000	1.000	1.000

Panel B: Pairwise model comparison tests for nested SDFs

	Restricted weighted χ^2 test			Wald test			Unrestricted weighted χ^2 test		
	Size			Size			Size		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.123	0.065	0.014	0.108	0.052	0.010	0.189	0.112	0.032
240	0.106	0.053	0.010	0.101	0.047	0.007	0.180	0.106	0.030
360	0.102	0.050	0.010	0.099	0.046	0.007	0.166	0.095	0.025
480	0.099	0.049	0.009	0.099	0.047	0.007	0.154	0.086	0.022
600	0.098	0.048	0.009	0.100	0.048	0.008	0.147	0.080	0.021

	Power			Power			Power		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%
120	0.337	0.198	0.046	0.469	0.334	0.130	0.311	0.188	0.050
240	0.505	0.343	0.112	0.647	0.524	0.284	0.451	0.311	0.115
360	0.634	0.474	0.195	0.760	0.656	0.421	0.563	0.419	0.183
480	0.732	0.585	0.285	0.836	0.751	0.535	0.657	0.518	0.256
600	0.807	0.681	0.377	0.888	0.821	0.634	0.737	0.610	0.333

Panel C: Multiple model comparison test for nested SDFs

Wald test

	Size			Power		
T	10%	5%	1%	10%	5%	1%
120	0.155	0.084	0.020	0.372	0.246	0.083
240	0.118	0.059	0.012	0.573	0.441	0.215
360	0.110	0.055	0.011	0.713	0.597	0.359
480	0.109	0.054	0.010	0.809	0.713	0.491
600	0.108	0.053	0.010	0.875	0.802	0.606

Table 3 (continued)

Model selection tests

Panel D: Pairwise tests of equality for overlapping SDFs

	Restricted weighted χ^2 test				Wald test				Unrestricted weighted χ^2 test		
	Size				Size				Size		
T	10%	5%	1%		10%	5%	1%		10%	5%	1%
120	0.091	0.042	0.007		0.109	0.053	0.010		0.101	0.051	0.010
240	0.091	0.043	0.007		0.099	0.048	0.008		0.108	0.057	0.013
360	0.093	0.044	0.007		0.098	0.047	0.009		0.108	0.058	0.013
480	0.094	0.045	0.008		0.099	0.048	0.009		0.106	0.056	0.013
600	0.095	0.045	0.008		0.100	0.049	0.009		0.105	0.054	0.012

	Power				Power				Power		
T	10%	5%	1%		10%	5%	1%		10%	5%	1%
120	0.649	0.560	0.359		0.776	0.654	0.374		0.418	0.327	0.171
240	0.888	0.857	0.758		0.980	0.957	0.851		0.722	0.663	0.528
360	0.959	0.948	0.914		0.999	0.997	0.982		0.870	0.838	0.760
480	0.984	0.980	0.968		1.000	1.000	0.998		0.941	0.925	0.883
600	0.993	0.992	0.987		1.000	1.000	1.000		0.973	0.964	0.944

Panel E: Pairwise ($p = 1$) and multiple ($p = 2$) model comparison tests for overlapping distinct SDFs

	$p = 1$							$p = 2$					
	Size			Power				Size			Power		
T	10%	5%	1%	10%	5%	1%		10%	5%	1%	10%	5%	1%
120	0.136	0.069	0.011	0.432	0.310	0.127		0.134	0.065	0.009	0.357	0.243	0.086
240	0.110	0.052	0.007	0.623	0.490	0.251		0.105	0.047	0.005	0.527	0.393	0.176
360	0.102	0.048	0.007	0.754	0.635	0.388		0.098	0.045	0.005	0.668	0.541	0.297
480	0.102	0.048	0.007	0.842	0.746	0.513		0.098	0.045	0.006	0.775	0.664	0.419
600	0.101	0.048	0.008	0.901	0.827	0.623		0.098	0.045	0.007	0.850	0.759	0.532

The table presents the empirical size and power of pairwise and multiple model comparison tests for strictly non-nested (Panel A), nested (Panels B and C) and overlapping (Panels D and E) models. In Panel A, we report simulation results for the weighted chi-squared and LM tests in parts (a) and (b) of Theorem 2, respectively. In Panel B, we report, in the order, results for the restricted weighted chi-squared test in part (a) of Theorem 3, the Wald test in part (b) of Theorem 3 and the unrestricted weighted chi-squared test in (55)–(56). Panel C is for the Wald test for multiple nested model comparison analyzed in Section 3.2. Panel D reports results for the restricted weighted chi-squared test in part (a) of Theorem 4, the Wald test in part (b) of Theorem 4 and the unrestricted weighted chi-squared test in (55)–(56). Finally, Panel E presents results for the pairwise ($p = 1$) and multiple ($p = 2$) model comparison tests in (46) and (74), respectively. We report results for different levels of significance (10%, 5% and 1% levels) and for different values of the number of time series observations (T) using 100,000 simulations, assuming that the factors and the returns are generated from a multivariate normal distribution.